MOBILITY AND SECURITY MANAGEMENT

7

MOBILITY AND SECURITY MANAGEMENT

The radio resource management plane, studied in the previous chapter, covers signalling aspects very specific to the radio interface of GSM. The communication management plane, which we will see in the next chapter, covers on the contrary functions of little GSM specificity, implemented in a way very close to their counterparts in ISDN. In between lies some other functions, not specific to the radio interface, but for the moment peculiar to cellular networks. These functions are not covered by the ISDN protocol specifications, and are grouped in an intermediate functional plane.

This modelling by exclusion results in a rather disparate set of functions in this plane. Two coherent groups stand out. The first one includes the functions rendered necessary by the movements of the subscriber. This management of the mobility has given the overall name of the functional plane. It includes two facets: how the mobile station deals with a changing environment, and how the infrastructure manages the subscriber location data, to enable efficient establishment of calls towards GSM subscribers.

If the allowance for the mobility of the subscribers and its automatic management is a fundamental service in a cellular network, it also raises some technical problems, which are amplified in GSM by the notion of roaming between networks operated independently. The study of the mobile station side will be centred on the way to choose the cell and the network when a choice exists. The management of the subscriber location data on the infrastructure side is the other facet of the problem.

and is concerned with getting and updating the location information needed to route calls toward a GSM subscriber who can move between cells or even networks.

The second group relates to the management of the security features of GSM, that is to say the protective measures against fraud or eavesdropping on the radio interface.

Both groups share common aspects when the implementation is looked at. They involve the same equipments, and they interact in some procedures. In both cases, the Subscriber Identity Module (SIM) and the Home Location Register (HLR) play an important role.

Some little related functions are added to these two groups, according to the modelling of the *Specifications*. They are dealt with rapidly in the last section of this chapter.

7.1. LOCATION MANAGEMENT

The mobility of the subscribers has major technical consequences in the infrastructure, but has also the important consequence that the service provided to a given subscriber changes as he moves, because of radio propagation (he may move out of coverage); because his subscription may be limited geographically; and because he may be served by different networks, providing different services.

In idle mode, the mobile station must choose *one* cell from which it expects to receive call attempts towards the subscriber. To this avail, the mobile station listens to the Paging and Access Grant Channel (PAGCH). It is said to be **camping on** this cell. The way the mobile station should choose which cell or which network it will camp on depends a lot on these service considerations. We will then present the mobility management functions starting with an overview, including the various factors influencing the service provided to the subscribers. Some are of administrative nature, another is propagation, and still others come from the system behaviour, such as congestion control. This will allow us to present the way the mobile station chooses (or helps the user to choose) between networks and between cells, while explaining the rationale behind these design choices. Only then will we tackle the infrastructure side, that is to say how the infrastructure keeps track of the subscribers' location.

7.1.1. THE FACTORS DETERMINING THE SERVICE

In the fixed telephone system, the service as it appears to a subscriber depends on which network the subscriber's telephone line is connected to, and hence on the location. For instance, the way a called number is entered, the price of the communication, the additional services that may be available, all depend on the location. Fixed network operators are working towards harmonisation, but the process is very lengthy. The consequences of these differences are however minor in the fixed system: in a system where the users move, the situation is quite different.

GSM has been designed to enable an international coverage, for instance users of the European GSM900 will be offered a European-wide system area. A subscriber will be able to get full access to the service from many countries in Europe with a single subscription. However, in order to adapt the system to various types of users, several levels of service may be offered on a geographical basis. For example, GSM operators may offer their customers different subscription choices, ranging from a (cheap) subscription limited to part of the country (regional subscription), to a (more expensive) subscription encompassing the whole area covered by GSM900 networks. This can go even further with SIM-roaming. If suitable agreements exist, the subscription may extend to networks of other types. Since the SIM interface is common to all GSM-based systems, the subscriber of a GSM900 network may obtain service from a DCS1800 network, and reciprocally, provided he uses an adapted mobile equipment.

In order to manage this flexibility, subscription in GSM is defined around the leading concept of PLMN ("Public Land Mobile Network"). This concept will be developed together with the description of the other administrative aspects.

Subscription on the one hand, and coverage limitations on the other, impact the services a user has access to when he moves. A first rough division distinguishes three levels:

- the "normal service", where a user can be called and can call, using all the services he has subscribed to (at least those the serving network can provide);
- the "limited service", where the only possibility left to the user is making emergency calls (typically in an area within coverage but not included in the subscription entitlement); and
- the "no service" case, typically when the user is completely out of coverage of any compatible network.

Let us see in more detail what determines the level of service given to a subscriber.

7.1.1.1. Administrative Aspects

The Notion of PLMN

The development of the Technical Specifications of GSM was concurrent with the organisation of a European-wide service by the would-be GSM operators. Most of the administrative features of the system have been deeply influenced by the context of European Telecommunications, and a number of these features have left their marks in the GSM Technical Specifications.

The European GSM system is divided into a number of separate operational networks, each being operated to a large extent independently from the others. Each of these networks is called a PLMN (Public Land Mobile Network—the term is much more generic than its specific usage in the *Specifications*). One of the restrictions, probably derived from the organisation of CEPT, is that the commercial coverage area of each PLMN is confined within the borders of one country. PLMNs of different countries may nevertheless overlap a little in border areas (radio waves have no respect for political borders). Most countries have several PLMNs, whose coverage areas overlap partly or completely: competition between operators is the rule of the game. Presently, licences for operating GSM900 or DCS1800 in Europe have been granted to typically two or even three operators per country. The operator may be a private company, a public company or an administration. The total number of European operators holding a GSM licence is in the order of 25 in 1992.

Roaming

It should be noted that the grouping of several operationally independent PLMNs in a single system open to roaming, in which the users can move and keep access to the service, is only possible if some conditions are met. First, the PLMNs must communicate between themselves. This requires standardised means of communication between PLMNs. Second, a subscriber must have a piece of equipment enabling him to access the different networks. As explained in the first chapter, GSM is designed to support MS-roaming, where the piece of equipment is the mobile station itself (thanks to the standardised GSM900 or DCS1800 radio interfaces), and moreover opens the door to SIM-roaming, where the piece of equipment is the SIM only.

The air interface and the inter-PLMN interface are the only standardisation requirements which are required to provide MS-roaming. For SIM-roaming, the standardisation of the inter-PLMN interface is still needed, but the only other requirement is a common interface between the SIM and the mobile equipment (SIM-ME interface).

Subscription

A GSM customer has a subscription relationship with a single PLMN. This specific PLMN is called the **home PLMN** of the subscriber. Service can be obtained from other PLMNs, depending among other conditions on subscription. In the *Specifications*, the term of visited PLMN (or VPLMN) is sometimes used to refer to a PLMN other than the home PLMN. In order to remove ambiguities, this term will be used in this book only when it is relevant to mention explicitly that the PLMN referred to is not the home PLMN of the subscriber. In other occasions, i.e., when the relationship with subscription is not relevant, the term "PLMN", or "serving PLMN" will be used.

Subscription information includes the set of services chosen by the user, as well as regional or international entitlements. Emergency call is the only service which is available anywhere in the system, whatever the subscription conditions. In fact, this service may even (in most PLMNs) be open to anonymous calls, i.e., calls for which no subscriber identity is mentioned. In this case, the SIM is not necessary, and a mobile equipment without SIM may be used for emergency calls. Access to normal service is of course a different matter: the home PLMN, the visited PLMN and subscription entitlements all have a role to play.

PLMN Accessibility

We will now look more in detail at the different conditions governing access to a given PLMN, taking into account roaming agreements and subscription limitations. To do this, we will follow a subscriber called Alan.

Access to the Home PLMN

Depending on his subscription, Alan can access normal service in the whole area covered by his home PLMN, or only in a part of it. The last case is referred to as a regional subscription. Presently, the home PLMN is the only one in which a regionally limited access is possible, on a subscription basis. There are no technical problems in doing otherwise,

although the way the mobile station selects cells in the phase 1 *Specifications* takes this restriction into account. This will be changed in phase 2, and regional subscription over several PLMNs may be offered, if the commercial interest is worth the complexity of the administrative steps.

According to the *Specifications*, the regional limits for subscription zones are constrained by the requirement that they must not include parts of VLR areas. A VLR relates to one or several MSCs, each controlling a number of cells. The coverage areas of all these cells form the VLR area. It would have been more flexible for operators to enable the management of regional subscription on a smaller basis, and indeed the restriction would be very easily alleviated by a change of the inter-location register protocol. In this case the subscription zones could consist of location areas (introduced in Chapter 1). It might even have been better on a cell per cell basis, because location areas should be designed so as to balance location updating traffic with paging traffic, and this traffic optimum would have been easier to reach without mixing in administrative aspects such as the boundaries of regional subscription. However, there are no simple means to go lower than the location area level without a large increase in the technical complexity. Because the list of cells composing a subscription area constantly evolves with network extension and reconfiguration, the mobile station cannot know this list beforehand. Only Alan's home PLMN is aware of this information. The mobile station has to learn in real time whether normal service can be provided or not in a given cell, by some enquiry means. This would be very costly in terms of signalling, unless done at the same time as location updating, and this is possible if and only if subscription area borders are also location area borders. This was the choice, and the location updating procedure has also the function of verifying subscription entitlement in the location area.

Access to PLMNs of the Same Country

The rules for roaming in PLMNs of the same country as the home PLMN are one of the aspects of the *Specifications* which had continued to evolve during the elaboration of the standard. GSM900 phase 1 and DCS1800 phase 1 offer different views, not to mention later phases.

In GSM900 phase 1, all PLMNs other than the home PLMN are treated on the same basis for selection, independently from their country. Access to them may be allowed or not depending among other conditions on subscription choices, but always on the basis of "everywhere" or "nowhere" within the PLMN coverage area.

When DCS1800 phase 1 was standardised, there was a strong will to introduce some more controlled form of competition between operators of the same country. Operators asked for a mechanism by which subscribers of other PLMNs of the same country could be tolerated in a (typically low-density) area where a single PLMN provides coverage, whereas these same subscribers could be barred in other (typically high-density) areas where several PLMNs provide coverage. Such a mechanism would allow each operator to install only part of the infrastructure needed for full coverage of low-traffic areas, whilst providing overall a full nation-wide service for customers. This mechanism was introduced in DCS1800 as early as phase 1, and is called "national roaming". It allows users to access parts of PLMNs in the same country, these parts being chosen by agreements between operators and not by subscription. Relevant areas may in fact evolve with the deployment of the networks, without the subscriber being directly aware of these changes.

As for regional subscription, national roaming is performed on a location area basis, for the same reasons. The mobile station must learn by location updating attempts which location areas are acceptable or not.

In future phases, the mechanism of national roaming will be part of both GSM900 and DCS1800.

PLMNs of Other Countries

Access to PLMNs in countries other than the home PLMN country is possible for subscribers entitled to it by subscription. If so, the access is possible in the whole PLMN area. This feature is referred to as "international roaming".

Since new PLMNs can be created any time, the mobile station does not keep a list of the subscribed-to PLMNs. It will learn if a PLMN accepts the subscriber or not by attempting location updatings.

Mobile Station Constraints

Because the service offered to Alan may depend on the PLMN used for access, an important feature of the mobile station is how it chooses, or how it helps Alan in choosing, the serving PLMN when several PLMNs are possible. Moreover, a number of operational details may change when the serving PLMN changes, such as cost and local dialling format. Therefore, Alan needs to be aware (and, if possible, in control of the choice) of the serving PLMN.

PLMN selection is one of the subjects which have been under discussion even after the freezing of phase 1. There are some differences between GSM900 and DCS1800, and there are much more important differences between phase 1 and phase 2. The whole issue arises from the conflict between two opposite aims: quick response of the mobile station to a change in the configuration of available PLMNs on one hand, and battery life on the other hand. Any solution reflects a compromise between these two ends.

Quick MS Response to PLMN Availability ...

If power consumption was not at stake, the solution would be to let the mobile station explore continuously the whole GSM900 (or DCS1800) spectrum for BCCHs. It would then detect as soon as possible any new PLMN, and take it into account in its selection algorithm.

... Versus Low Power Consumption

The monitoring of the radio environment for beacon frequencies is an operation that costs power consumption. Now battery life is a very important aspect of a handheld mobile station: no subscriber would accept such a device if refuelling was needed every second hour. Therefore a trend is to try to limit as far as possible the search for neighbouring cells done by the mobile station.

The scheme adopted for GSM900 phase 1 is an extreme example of this method: the mobile station in normal service only monitors the cells in the same PLMN and in the neighbourhood of its serving cell. To this end, the serving cell broadcasts the list of the beacon frequencies used by neighbouring cells. This scheme is very efficient for limiting power consumption, but not for finding alternative PLMNs when the user moves into an overlapping area. The resulting behaviour is not optimum as seen by the PLMN operators: they would like the mobile station to make the right choice, in particular when the home PLMN becomes available for a mobile station who was being served by another PLMN.

In the cases where the list of frequencies used by neighbouring cells is not available, or more generally when a PLMN selection needs to take place, the mobile station has to search the whole spectrum.

7.1.1.2. Radio Considerations

Administrative considerations are not, by far, the only factor determining the service a given cell may provide to a given subscriber. A very important aspect is that the transmission between the base station

and the mobile station must offer a good quality. Radio propagation considerations must then be within the criterion for the choice of the serving cell.

In idle mode, the only thing the mobile station has to do is to listen to the information broadcast by the cell it camps on (including the paging requests). If we were to take only this into account, then the best cell to choose would be simply the cell which has the best reception level or quality. But the rule in GSM is that when a mobile station wants to exchange information with the network, e.g., to set-up a call at the user's request, or to answer a paging, it must do it in the cell it is camping on. It could have been different: one could imagine that the mobile station selects the cell to communicate with at the very last moment. Of course, nothing precludes the network to perform a handover very soon after receiving a call request. This mechanism, called "directed retry", is however far from systematic in GSM networks: in the general case, the call will stay for some time in the same cell as was selected by the mechanism in idle mode.

Because of this choice, the camped-on cell should also be as close as possible to the best cell in which a potential connection will be set up. As a consequence, the quality of reception by the mobile station must not be the only parameter taken into account, but also the quality of reception by the base station. This cannot be measured directly in idle mode, since the mobile station is not transmitting, but this can be derived from reception measurements and from the maximum power the mobile station can use for transmission.

A criterion used to choose a cell in idle mode combines the reception level of the mobile station on the beacon frequency, the maximum transmission power of the mobile station, and several parameters depending on the cell (and broadcast on the BCCH). The exact algorithm is described later on (see the description of the C1 criterion, page 453).

7.1.1.3. System Load Control

Congestion is a risk which exists in any telecommunication system. Mobility changes slightly the bases of the problem. The traffic variations are of bigger amplitude, since they come not only from the change of traffic per subscriber, but also from the movements of the subscribers. For instance a sport event may see a huge concentration of mobile subscribers in one small place at the same moment. Another point is that, because subscribers are not physically linked to a cell, they may "move" from a congested cell to another if the second one can provide the

service. Another aspect to look at in connection with cell selection is then the way in which a network can control the traffic distribution among cells. Two mechanisms exist, one of which impacts cell selection.

The network can bar completely a cell against access by all normal subscribers. This "barred" status is indicated in the information broadcast by the cell, in the *CELL_BAR_ACCESS* flag described in Chapter 6. Such a mechanism is used when a BTS is unable to operate properly, e.g., for maintenance purposes. It may also prove handy when the operator sets up new cells and performs tests on these cells before opening them for normal operation. Test mobiles (which ignore the "barred" status, and do not then conform to the *Specifications*) are able to establish connections with such cells for test purposes. Another potential application of cell barring concerns cells restricted to handover access.

Cell barring is an all-or nothing control mode, which must be taken into account by the mobile stations for cell selection in idle mode. GSM also includes a subtler mechanism, called the access class mechanism, which allows selective access of certain mobile stations to certain cells. Its purpose is to cope with abnormally high traffic load or emergency situations, but it does not influence cell selection. For example, a mobile station may perfectly select and camp on a cell which will not at this very moment accept a connection request from this particular mobile station (even for location updating purposes) because of a temporary high load. Allowing such a cell to be nevertheless selected avoids the congestion situation to spread in neighbour cells. This is an example of a choice in the specifications where a global optimum, evaluated on several cells, is favoured against the local improvement. The topic of access class has already been developed in the appropriate place, in the Radio Resource management chapter.

7.1.1.4. Paging and Location Areas

Finally, an important point to take into consideration for the PLMN and cell selection is that the network must be able to route calls toward the subscriber. The infrastructure must know some minimum information concerning the location of the subscriber to do so. This information can be provided only by the mobile station, and the service provided to the user depends on the consistency between the location currently assessed by the infrastructure and the cell chosen by the mobile station. It is then necessary to look in general terms at how the infrastructure deals with calls toward GSM subscribers.

In order to avoid a waste of signalling, the system is so designed that a subscriber is only looked for (paged) in a few cells of the system

when a call toward him has to be established. Cells are grouped in location areas, and a mobile station is typically paged only in the cells of one location area when an incoming call arrives (see the basic concepts of location management, in Chapter 1). Therefore, the mobile station must inform the system of the location area in which the subscriber should be paged. It does so by a location updating procedure. The network, on the other hand, must store the present location area of each subscriber: this storage is inside location registers as will be detailed together with the description of the location updating procedure. Each change of location area puts an extra load, not only on the radio path, but also on the infrastructure equipments: the cell selection mechanism therefore includes some features to limit the number of location updatings.

Location Areas

For obvious technical reasons, the *Specifications* impose that each location area be a subset of the cells of a single PLMN. In fact, because of the way a mobile terminating call is routed in the network, a location area must include cells managed by a single MSC (see figure 7.1). This restriction to an MSC could have been avoided, but only at the price of complex procedures. Within these constraints, the operator has complete

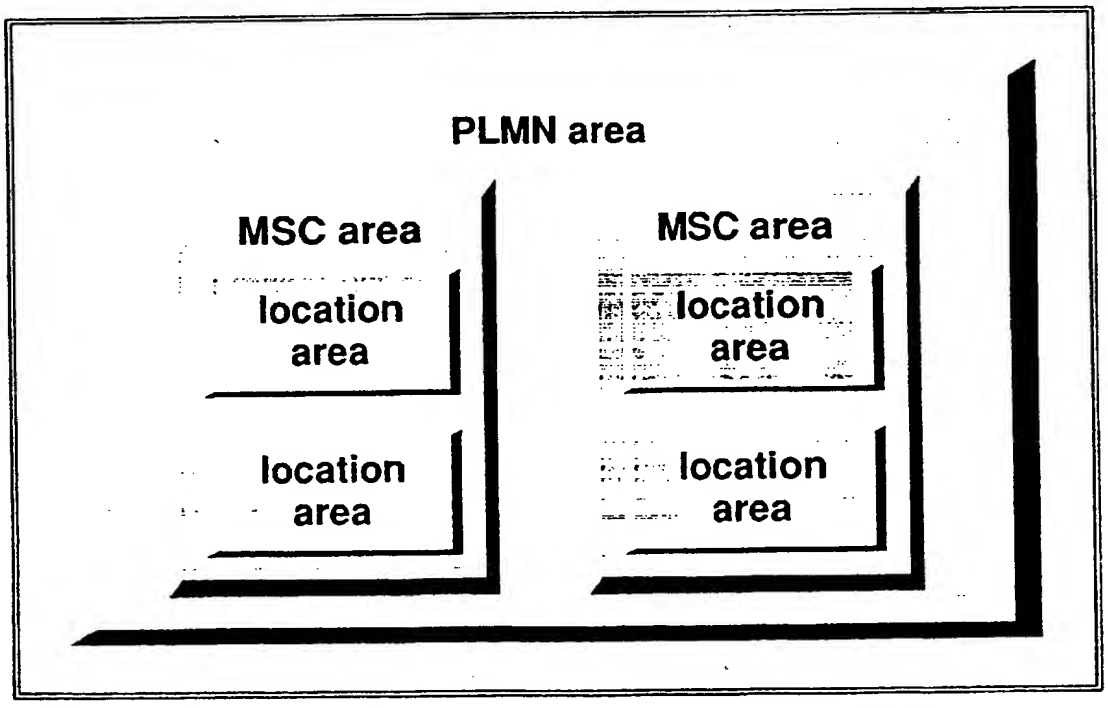


Figure 7.1 – Location area vs. MSC and PLMN areas

A location area may only contain cells of a single MSC, in a single PLMN.

freedom to allocate cells to location areas. The goal of this operation is to minimise resource consumption, taking into account the signalling load on the radio path (both from paging and location updating) as well as the processing load of the equipments.

Location Updating

If a mobile station wants to obtain normal service from a cell, and in particular to receive calls, it must make sure its subscriber (represented by the SIM) is registered in the location area of this cell. The registration state of the subscriber, except in network failure cases or after some very long inactivity period, can only be changed at the initiative of the mobile station. The outcome of the last registration attempt is stored in the SIM, as well as the identity of the location area. If this storage indicates success, normal service is assumed automatically if the mobile station camps on a cell of the same location area.

The situation is different if the mobile station is switched on in a different location area than the one where it was last successfully registered; when the mobile station moves into a place where a cell from another location area is better suited; or when the mobile station tries to get normal service in a different PLMN. In all these cases, the mobile station must attempt to register the subscriber by performing a location updating procedure prior to camping on the cell in normal service.

Status after Location Updating

Several outcomes of location updating are possible. As we have seen, the location updating procedure does more than just tell the network where the subscriber is. It also provides the network with a mechanism to tell the mobile station whether the cell can grant normal service, taking into account subscription limitations, national roaming restrictions, and so on. The best outcome of the procedure is when the procedure is run correctly and the network indicates that normal service is possible. The subscriber is then registered correctly in the network, and the mobile station is allowed to camp on the cell in normal service mode.

Several outcomes are possible when the registration is not successful. We must distinguish the cases where the procedure is run correctly (i.e., the network provides a meaningful answer, possibly denying normal service) from the others (i.e., the network does not answer or answers that it cannot give a meaningful answer). It is worth mentioning that the *Specifications* put all these cases together in an

"abnormal cases" category, but there is quite a difference between no answer and a bad answer!

Meaningful negative outcomes can be of three kinds, corresponding to the different limitations presented in the administrative aspects section:

- the cell may belong to a PLMN not supported by the subscription. Mechanisms are specified ("forbidden PLMNs" list, see further on) so that cells of this PLMN will not be tried anew, except on explicit request from the user. A new PLMN (and hence a new cell) will be looked for if normal service is sought.
- the cell can be in a location area not suitable because of regional subscription. The rule is then that the mobile station must stay in the cell (and in the home PLMN), but only limited service can be provided to the user. There are some reasons for this choice: because of the implicit assumption that, when a subscriber is only entitled to regional subscription in the home PLMN, access to other PLMNs of the same country is most probably not allowed by the operator: therefore looking for other PLMNs is not essential. However, this choice leads to a peculiar behaviour of the mobile station when the subscription covers a part of the home PLMN *and* foreign PLMNs. In such cases, the mobile station will look for these other PLMNs only when leaving the coverage of the home PLMN, not when leaving the subscribed-to part of it.
- in DCS1800, the cell can be in a location area which does not accept roamers of another PLMN of the same country. As with non-subscribed-to PLMNs, mechanisms are included to prevent further attempts in the cells of the same location area. In addition, the mobile station will look immediately to see if the home PLMN is available.

An important point is that in some cases (if reject comes from the home PLMN or from a PLMN which is connected to the home PLMN), the subscriber is effectively deregistered in the HLR: it is marked in the network as not being in a state where calls to his destination can be established. Such calls are either directed to an announcement, or forwarded to another destination if the subscriber has activated call forwarding unconditional or call forwarding on not reachable. In the other cases, the subscriber keeps the same HLR registration state he had before the attempt. However he is considered as deregistered by the mobile station, since there is no indication from the network about what happened with the HLR state.

When the mobile station receives one of the negative answers listed above, it may look for other PLMNs. If none of the found PLMNs are able to provide normal service, the mobile station goes to the limited service state.

Cases of procedural failure are even more complex. Details can be found in the *Specifications* and will not be dealt with here. Basically, the mobile station will not try other cells (unless the radio criteria so determine) despite the lack of knowledge about the service that can be provided to the user: this is an important point. There is a first phase during which the mobile station will try several times to get an answer from the network. During this phase, the mobile station will behave as if it was granted the same service as before. If all attempts fail during this phase, the mobile station enters a special state, in which it does not assume any registration state in the network. It tries now and then a location updating procedure, to get out of this unsatisfying state. The mobile station does not reject a request from the user, but this request triggers the mobile station to start a location updating, and the mobile station will eventually reject the user's request if the network still does not answer positively.

7.1.2. CELL AND PLMN SELECTION

The cell and PLMN selection mechanisms are almost totally specified in the *Specifications*. We have seen in the previous section the different factors to take into account, as well as some general mechanisms. Let us see now the detailed process.

7.1.2.1. PLMN Selection

Though the PLMN selection process (as well as cell selection) affects only the mobile station, it is specified in a fair amount of detail in the *Specifications*. One of the reason is that the SIM intervenes, and the SIM-ME interface has to be specified. Other reasons are to harmonise somewhat the behaviour of the mobile station, so that it can be predictable when a user changes his equipment, and also to avoid the possibility that some implementations bias the choice of PLMN and thus the competition between operators.

However, the *Specifications* are not totally constraining, and have to be completed for some marginal cases by the mobile station manufacturers. We will mention these latitudes, with some of the possibilities. What the mobile station really has to do is to select a cell.

There is however an important difference between the selection of the serving PLMN, and the selection of the serving cell. The first is under control of the user, whereas the second is fully automatic. The PLMN selection is important for the user when the service finally obtained is normal service. In the limited service case, the choice is of little importance, and in the case where no PLMN can be found the selection of a PLMN is of no immediate application. We will look at the three cases in turn, before summarising the user's view.

The Normal Case

If already in normal service mode the mobile station only looks for the cells in the serving PLMN, independently from other PLMNs. A change of PLMN can occur at only two occasions: when the user decides so, and when the mobile station finds out that the serving PLMN can no longer provide normal service (e.g., because the mobile station is leaving the PLMN coverage area). In those cases, the mobile station will search for cells in the whole spectrum, to find which PLMNs cover the location. Access to some of these PLMNs is then tried, according to the PLMN selection method. Two methods are provided for the choice of the PLMN to try, the choice being left to the user: the **manual mode** and the **automatic mode**.

Several aspects are common to both modes. For example, the home PLMN is set as the PLMN to try at switch-on, independently from previous history (even if the mobile station knows, from the SIM, that the subscriber is currently registered in another PLMN). Another common aspect between the manual and automatic modes is the "*forbidden PLMNs* list—the first one in a series of 3 PLMN lists to juggle with! The actual list of PLMNs accessible to a subscriber according to his subscription may change as new PLMNs are opened (or closed!) for service. The list of these PLMNs cannot be stored once and for all (for instance in the SIM). Instead, it has been chosen to build dynamically a list of PLMNs which are *not* accessible according to subscription. This list is updated according to the result of access attempts performed by the mobile station, and is stored in a non-volatile memory in the SIM (and hence is not lost when the mobile station is switched off). The list is limited by the *Specifications* to 4 entries, but nothing precludes a longer list to be stored in the ME (therefore possibly lost at switch-off).

This so-called *forbidden PLMNs* list includes PLMNs which are not subscribed to. They are not really "forbidden", since they could be used for emergency calls, and in manual mode the user is perfectly authorised to select one of them. When a mobile station attempts to update its location in a PLMN to which the user's subscription does not

authorise access. the network will tell the mobile station so, and the PLMN identity will be put in the list, ejecting if needed the oldest entry. The *forbidden PLMNs* list is used for PLMN selection both in manual and automatic mode, though in different ways.

Another list, this time stored in the mobile equipment (not in the SIM) intervenes a little in the PLMN selection process in automatic mode. It contains the identities of the location areas which have rejected access before because of national roaming limitations. Cells belonging to the location areas in this list can no longer be candidates for selection, and a PLMN for which all the found cells are in this category cannot be tried in automatic mode. This list is filled as a result of location updating rejections with a suitable cause (but the PLMN is not put in the *forbidden PLMN* list), and is erased when the mobile equipment is switched off, or the SIM removed.

Manual Mode

In manual mode the list of PLMNs the mobile station has found as potential candidates for providing normal service is presented to the user, whether or not they are in the *forbidden PLMNs* list. This list of *found PLMNs* is displayed using explicit names (such as D1-Telekom or D2 Privat for the German PLMNs, DK TDK-Mobil or DK Sonofon for the Danish PLMNs, etc.) and with an explicit mention telling the user whether the PLMN is in the *forbidden PLMNs* list or not. Normally the user then chooses one of the PLMNs of the *found PLMNs* list as the selected PLMN.

The explicit names of the PLMN cannot be found in the *Specifications*, though they are part of the mobile station specifications. The list is distributed by the GSM MoU, and regularly updated. There is no requirement on the update of already manufactured (and sold) mobile stations, or on the way in which they should visualise the names of new PLMNs who have been introduced meanwhile. Typically, if the knowledge of the network commercial name is unknown, the mobile station will display the country initial together with the numerical network code, for instance "DK Network 05".

The user can choose any PLMN in the *found PLMNs* list, even if it is also part of the *forbidden PLMNs* list. This possibility might be useful after a change of subscription category: for example, if Alan requests his subscription to be changed from "home PLMN only" to "all PLMNs" while he roams abroad, he might want to force his mobile station to attempt location updating on a network previously known as "forbidden", in order to unlock the situation.

Since, at switch-on time, the mobile first looks for the home PLMN and does not (in manual mode) take into account the PLMN the subscriber is registered in, some situations can become quite irksome when the user stays in a foreign country for some time. Automatic mode is in such cases more appropriate.

Once the user has selected a PLMN, the mobile station will attempt to get normal service on this PLMN. Several outcomes are possible, as we have seen. It may well succeed: fine! The PLMN may also be indicated as not allowed by subscription, and will therefore join the *forbidden PLMNs* list. Unfortunately, the other cases are not so simple, and not always clearly specified. In manual mode, these cases can be treated generally by letting the mobile station do its best to get normal service (possibly by trying several cells in different areas). After some time, the PLMN list is presented again, including the PLMN on which the attempt failed: it is up to the user to choose another one... If everything fails, the mobile station should at some stage reach the "limited service" described below.

Automatic Mode

In automatic mode, the mobile station will choose which PLMNs to try all by itself. The automatic mode is based on the existence of another list of PLMNs, the *preferred PLMNs* list, which is stored in a non-volatile memory in the SIM. This list, capable of holding at least 8 entries, includes a number of PLMN identities in order of preference and is under the control of the user. The most preferred is usually the home PLMN, but it is nowhere specified whether the home PLMN should appear at first rank in the list or not, whether the user may choose to have a list with a different PLMN as "top of the list" and what happens in this case. The list may originally be filled in by the home PLMN operator, during the SIM personalisation process. It can afterwards be modified at will by the user through a mechanism to be specified by the mobile station manufacturer. No automatic modification of the list can take place.

When a PLMN selection takes place in automatic mode, the PLMNs are tried starting with the first PLMN in the list of *preferred PLMNs* which is in the list of *found PLMNs* and not in the *forbidden PLMNs* list. The treatment of one of the failure cases is clear: if the PLMN is indicated as not allowed by subscription, its identity is put in the *forbidden PLMNs* list and then cannot be automatically selected again, except in the (rare) case where more than 4 PLMNs are found, none being allowed by subscription: in that case, the mobile station may rewrite the *forbidden PLMNs* list in a cyclic manner! In the other failing

cases, the mobile station must try the other possible PLMNs, in order of preference for those in the *preferred PLMNs* list. It is not clear what is the status of PLMNs found by the mobile station but not included in the *preferred PLMNs* list; the most literal interpretation is that they cannot be selected. This raises different problems, for instance in the case where the list is empty. It seems more logical for the mobile station to also include them in the choice process, according to some ordering rule, for instance randomly. The situation where none of the *found PLMNs* can provide normal service is not clearly specified in the *Specifications* either, but it seems logical that the mobile station should go to the limited service state in that case. The mobile station should keep some record of its attempts to avoid a deadly loop in which it will stubbornly attempt to get service from a PLMN without success when another would grant normal service.

As in manual mode, the user can force PLMN selection at any moment. The process is then the same as described above when the serving PLMN ceases to provide normal service, except that the serving PLMN is in the list of *found PLMNs*. The mobile station will then stay as the selected PLMN, except if a PLMN with higher preference rating has appeared since the last PLMN selection.

The Limited Service Case

In limited service mode, the only service available is emergency call. The theory is that all PLMNs can provide this service. The purpose of selecting a PLMN is then fairly limited. It could however prove useful for two main reasons: in border areas, the user might want to control the choice of the country where a potential emergency call would be routed; besides, the user could choose a PLMN not for immediate use, but to be the one used for normal service as soon as the mobile station can find a cell of this PLMN.

The *Specifications* specify that the choice of the cell in limited service mode is done generally independently from the PLMN or the location area. This general rule admits a noticeable exception, when the home PLMN is available for limited service, and no other PLMN is available for normal service provision.

In parallel, the mobile station monitors continuously (though possibly at a slow rate to save its battery) the 30 strongest carriers it receives for new PLMNs. The phase 1 *Specifications* leave open the behaviour of the mobile station, especially in manual mode, when a new PLMN is found. In automatic mode, the mobile station will perform a PLMN selection and try to get normal service on a newly found PLMN which is not in the *forbidden PLMNs* list (and for DCS1800 if the

location area is not forbidden either). In manual mode, one behaviour would be to ask the user every time a PLMN pops up: this could however lead to annoying situations in border areas where PLMNs may appear and disappear very frequently. Another approach for the manual mode could be to stay in limited service mode as long as some PLMN chosen by the user stays unavailable, whatever other PLMNs pop up, which is not very much more satisfactory. It is hoped that mobile station manufacturers will find a satisfying compromise.

When the home PLMN is available, the limited service mode can happen only when the user has a regional subscription and is in a part of the home PLMN area where he is not entitled to normal service. In this case, the *Specifications* impose a cell selection mechanism identical to the normal service state, i.e., limited to the cells of the PLMN. The mobile station tries each new location area of the home PLMN that pops up so as to find one granting normal service, but will go on another PLMN (on which the subscriber may be entitled to service) only when leaving the home PLMN coverage area, or when asked by the user. The mobile station does not store any list of the location areas in which the user is or is not entitled to normal service. Protection against trying again and again the same location area is obtained by the hysteresis mechanism for the change of location area, as in normal service mode (see further on).

The No Service Case

If no cell can be found at all, the 124 carriers in the band (374 for DCS1800) must be monitored, not just the 30 strongest carriers found as in the other cases. Otherwise, the mobile station behaves as in the limited service case, except that it cannot accept even emergency calls.

What the Users see

The mobile station must indicate the service state to the user. Different levels of precision seem allowed. Typically the information provided by the mobile station to the user distinguishes the normal service, the limited service and the no service cases. We will see further along that in cases of signalling failure an additional state exists, but it is not necessarily indicated to the user. In addition, the mobile station must be able to indicate to the user, possibly on request, the serving PLMN when in normal service state. The PLMN identity is given in clear text when possible, including country initials and network name.

For the users, PLMN selection is first a choice between manual mode and automatic mode. Some control must be provided to allow the

user to know whether the mode is set to automatic or manual, and to change the mode.

In automatic mode, the user usually does not intervene in the selection mechanism. He is however able to direct the process by two actions: he controls the *preferred PLMNs* list and he can ask for a forced PLMN selection at any time, in which case the mobile station will search the whole spectrum and select a PLMN, possibly the same as before. In order to enable the user to control the *preferred PLMNs* list, the mobile station must provide commands to display the list and to edit it. Whether a modification of this list is taken immediately into account (e.g., by a forced PLMN selection) or not is left open to mobile station manufacturers.

In manual mode, the user has a total control, but is solicited in each case, maybe too often in some situations. He is asked for a PLMN choice in several instances. This happens after switch-on (or SIM activation) if the home PLMN is not available. This happens also when the mobile station moves out of coverage of the serving PLMN. In addition, this may happen in DCS1800 when the mobile station was in a visited PLMN of the home country, and cannot find any more a cell that accepts it. In this case, a forced PLMN selection happens, and hence the user will be asked to choose a new PLMN. Finally, a prompt to select a PLMN may appear when not in normal service mode, and a new PLMN is found. As in automatic mode, the user can in addition force a PLMN selection at any time.

Whatever the triggering event, the list of all found PLMNs is presented to the user, including those in the *forbidden PLMNs* list, though with a distinguishing mark. The order of presentation was a topic raising some discussion between operators, since it was felt that this order may influence the user's choice. The issue was settled on the specification that the order should be random.

7.1.2.2. Cell Selection

As explained in the requirements, only cells can be chosen with which transmission will be a priori of at least minimum performance, and the cell choice should aim at maximising the transmission quality. The radio criteria therefore play the foremost part in cell selection. Thus, before describing the actual cell selection algorithm, we will first study them.

Radio Criteria

In order to maximise transmission quality, a criterion has been defined, which takes into account the level of the signal received by the mobile station on the beacon frequency, the maximum transmission power of the mobile station and some parameters specific to the cell. This criterion is named *C1*. (There is no *C2* in the phase 1 *Specifications*, but it will appear in phase 2.)

Description of the C1 *Criterion*

C1 is defined as follows:

$$C1 := (A - \text{Max.}(B,0))$$

A := Received Level Average – *p1*

B := *p2* – Maximum RF power of the mobile station

(all values expressed in dB)

The two parameters *p1* and *p2*—called respectively *RXLEV_ACCESS_MIN* and *MX_TXPWR_MAX_CCH* in the *Specifications*—are broadcast by the cell. The first one can take a value between −110 dBm and −48 dBm, the second between 13 dBm and 43 dBm in GSM900 (in DCS1800, the range is different). The second parameter has another independent usage: it represents the maximum transmission power a mobile station is allowed to use on the RACH. Because the range of the mobile station maximum transmission power is 29 to 43 dBm, only this sub-range of the second parameter is useful, whether for *C1* or as a maximum transmission power on the RACH.

C1 is used as follows. When looking for cells, either when looking for neighbour cells in normal service mode, or when searching PLMNs, only cells of positive *C1* (calculated from the *p1* and *p2* broadcast by each cell) are taken into account. When a choice between cells has to be made, the cell of best *C1* is chosen among those equivalent for other criteria. As a consequence, *C1* determines two things:

- the coverage limit of each cell taken in isolation, in the sense that outside the area where C1 is positive, the cell does not exist for the mobile stations;
- the boundary between two adjacent cells for selection in idle mode, determined as the locus where *C1*=*C1'*. The boundaries

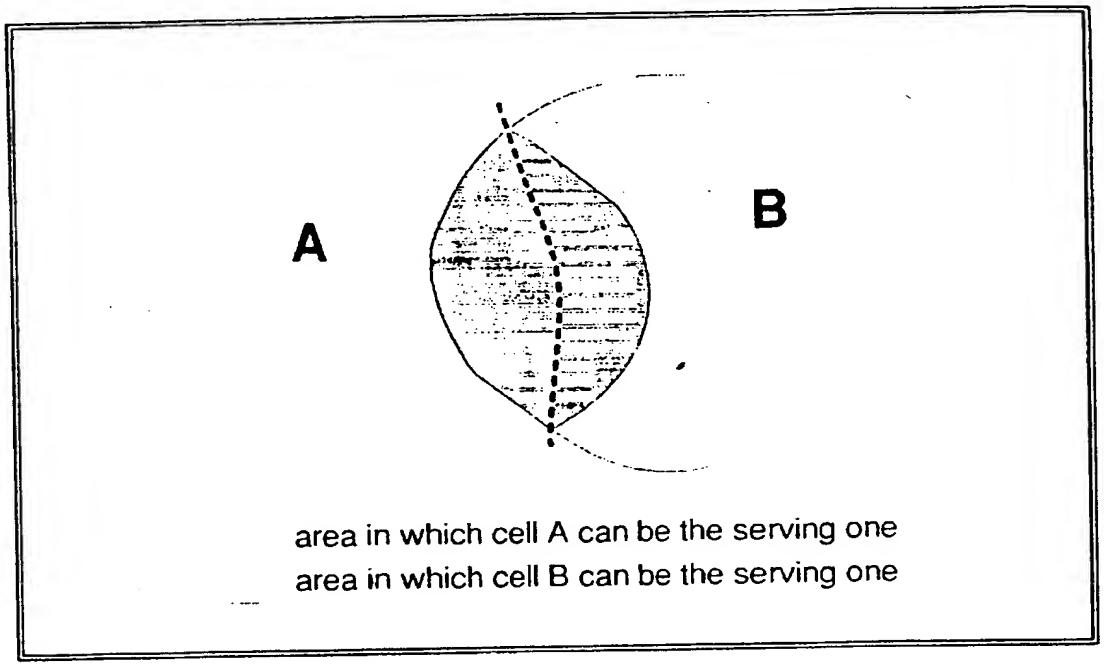


Figure 7.2 – Cell boundaries according to $C1$

The figure shows the boundaries of two cells A and B, according to the values $C1_A$ and $C1_B$ of the $C1$ cell selection criterion. The dashed line is the locus where $C1_A = C1_B$. Since $C1$ depends on the mobile station maximum power, these boundaries differ from one mobile station class to another.

with all adjacent cells determine a second cell limit, usually inside the area delimited by $C1$=0.

Figure 7.2 shows an example of two cells, with their $C1$=0 limits, and the line of equal $C1$'s. Two points are important to keep in mind with these limits. Because the maximum transmission power of the mobile station intervenes in $C1$, the limits are different for different mobile station classes. Second, that other cell limits exist, the ones determined by the selection of the cell for handover. It is up to the operator to choose $p1$ and $p2$ to obtain the correct compromise between cells boundaries, traffic and quality of transmission for the different classes of mobile stations, as well as consistency with the handover algorithms and parameters.

Criteria Other than C1

Because of its radio-electric nature, $C1$ varies quickly and to some extent randomly around a mean value depending on the location and on the movement of the mobile station. This means that the mobile station would often change between cells if $C1$ were the only selection criterion.

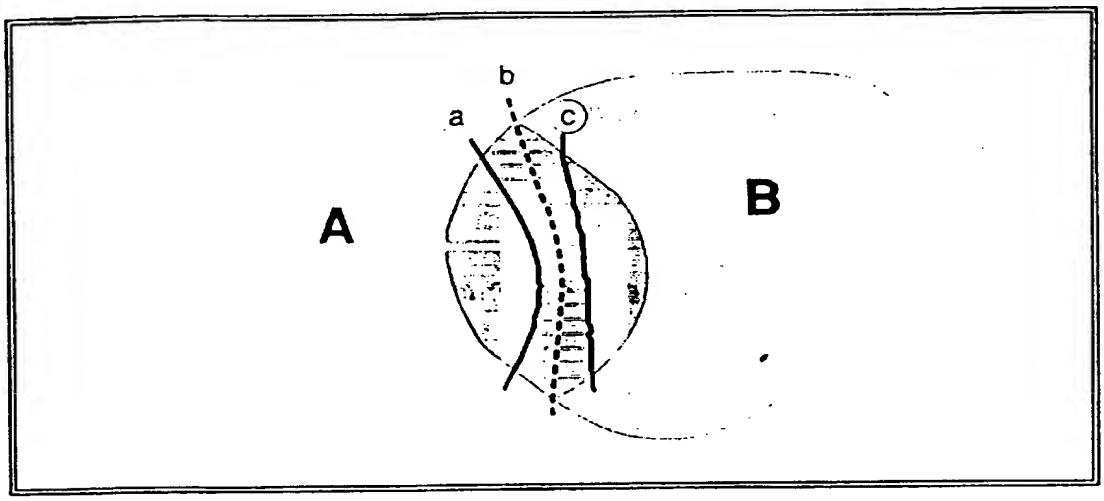


Figure 7.3 – Impact of *CELL_RESELECT_HYSTERESIS* on cell boundaries

The boundary between two cells A and B differs depending on the requirement for a mobile station to perform location updating:
(a) A and B belong to different location areas, and the mobile station is registered in the location area of cell B;
(b) A and B are in the same location area;
(c) A and B belong to different location areas, and the mobile station is registered in the location area of A.

This is acceptable if the cells are equivalent, but not otherwise, for instance between two cells belonging to different location areas. This is why the comparison between *C1's* is modified to be biased against the cells on which camping on must be preceded by a location updating. This is obtained by a handicap, added to the *C1* of these cells. The value of this handicap is not fixed, and is broadcast by each cell. It is called *CELL_RESELECT_HYSTERESIS*, though it is not strictly speaking a hysteresis value (the real hysteresis is the sum of the values in the two cells).

A consequence of this specification is that the boundaries between adjacent cells belonging to different location areas is not the same whether the mobile station goes from one cell to the other or the contrary. The hysteresis in terms of received level is transformed by the variations of propagation into a geographical hysteresis. Taking the same cell configuration example as before, figure 7.3 shows the cell boundaries as obtained when such handicaps are used. We have three different boundaries for a given mobile station class: one for mobile stations going from cell A to cell B, one for mobile stations going from cell B to cell A, and a third for mobile stations which do not apply the bias (e.g., foreigners not entitled in the PLMN by subscription). Mobile stations located in the area between lines a) and c) can be attached to either of the cells, depending on the direction they come from.

The Cell Selection Algorithm

The different requirements the cell selection algorithm has to meet have been detailed in the previous paragraphs. This algorithm is specified in the *Specifications*, and we will present it, as a summary of the points seen so far.

The aim of cell selection can be summarised as follows: in order to get normal service, the mobile station must camp on one of the cells fulfilling the following conditions:

- a SIM must be inserted, and the corresponding subscriber registered in the location area the cell belongs to;
- criterion *C1* for the cell must be higher than 0;
- the cell must nof be barred.

Among the cells which comply to these three requirements, the chosen cell must fulfil the two additional conditions:

- the cell's *C1* must be higher than the *C1* of any other cell found by the mobile station in the same location area;
- the cell's *C1* must be higher than the *C1* of any other cell found by the mobile station in different location areas of the same PLMN, corrected by the applicable handicap factor.

As can be noted, potentially better cells in PLMNs other than the one the mobile station is registered in are not taken into account. This corresponds to the decision that PLMN selection is triggered only by the user, or when the mobile station leaves the coverage of the selected PLMN.

Cell Selection in Normal Service State

Having this goal in mind, the behaviour of the mobile station can be easily derived. Let us start with the mobile station being in normal service state (hopefully the likeliest state). In this state, the mobile station receives a list of frequencies (broadcast by the serving cell) indicating where to look for the beacon channels of the neighbouring cells of the same PLMN. The mobile station must then find these beacon channels one by one and get their synchronisation information in order to decode some of the broadcast information they carry; this information enables the mobile station to check the PLMN, to know if the cell is barred or not and to obtain the identity of the location area the cell belongs to, as well as to get the various radio parameters so as to compute *C1*. Only

acceptable cells (i.e., non-barred and of positive *C1*) of the right PLMN will be taken into account. The mobile station can then compare the *C1* of these cells with the *C1* of the current cell. All this process takes place in parallel with the periodic reception of the paging channel on the current cell.

If the mobile station finds a better cell in the same location area, it changes to this cell and goes on with the process of listening to the paging channel (of the new cell) while monitoring the beacon channels in the new list. However, if the mobile station finds a better cell in a different location area of the same PLMN, having taken the location updating bias into account, it changes to this cell. At this very moment, the mobile station is no more—strictly speaking—in the normal service state: calls to it will in general not reach their destination. The mobile station tries immediately a location updating procedure to warn the network of its new location. Most of the time, the mobile station is granted normal service by the network, and is back to the normal service state in the new location area after a few seconds.

Cell Selection at Switch-on Time

Let us now describe other cases. An important one is the initialisation: how to get normal service in the first place after switch-on. The first PLMN to try (if found) is the home PLMN, according to the *Specifications*. The mobile station must search for non-barred and C1-positive cells within this PLMN. Without any information, the mobile station has to search the whole spectrum for beacon channels. This can be a lengthy operation, in particular in DCS1800 where 374 frequencies are supported.

The *Specifications* include a mechanism to help the mobile station in such conditions: a list of frequencies to look for can be stored in a non-volatile memory (in the SIM). The *Specifications* are however not clear about which frequencies the mobile station should put in that list. To be consistent with the imposed search for the home PLMN at switch-on time, it should be the neighbouring cell frequencies broadcast by the last cell of the home PLMN on which the mobile station camped (whether for normal service or not, and whether or not the mobile station was registered on other PLMNs meanwhile).

Whatever the case, the search results in a list of acceptable cells, with their *C1*. If this list is not empty, the mobile station chooses the cell of best *C1*. Also stored in a non-volatile memory is the identity of the location area (if any) in which the mobile station knows it is registered. If the chosen cell belongs to that particular location area, the mobile station goes immediately to normal service, possibly indicating its

presence to the network (see the description of the "IMSI attach/detach" mechanism, page 474). Otherwise, if the chosen cell is in a different location area, the mobile station camps on the cell and starts a location updating procedure immediately.

If no acceptable cell of the home PLMN is found, the mobile station acts as if it were leaving the home PLMN coverage area.

Cell Selection at PLMN Change

Another special case is when the mobile station has to look for the available PLMNs, for instance because the mobile station has moved out of the coverage of the PLMN previously providing service, or because of a forced PLMN selection by the user. The process is similar to the switch-on case, except that the mobile station has no information whatsoever about which frequencies to search: it must search the whole spectrum. The mobile station proceeds in two steps: first, it searches all GSM (or DCS1800) carriers, then it selects the 30 strongest ones to obtain the information they broadcast: the PLMN to which they belong, whether they are barred or not for access, and the parameters controlling *C1*. The mobile station can then establish the list of the acceptable PLMNs, that is to say those for which it has found at least one acceptable cell. This will result in a *found PLMN* list as described some pages before. When one of the PLMNs in the list is chosen, the mobile station will access the acceptable cell of best *C1* within those previously found in this PLMN and request a location updating.

Cell Selection in Limited Service Mode

Now we have to address the cases where normal service cannot be granted, but limited service is possible. This happens when the subscriber is not entitled to normal service in any of the found PLMNs. If the home PLMN is acceptable, i.e., if access to the home PLMN is locally prevented by subscription rather than radio propagation, the cell selection is the same as for the normal service state. Otherwise, the mobile station selects the acceptable cell of best *C1*, irrespective of the PLMN or the location area of the cells, and hence without applying the location updating bias. The mobile station searches continuously the whole spectrum for new cells, in order to find an acceptable PLMN as soon as possible. When such a PLMN is found, it may be selected and then the mobile station will try to get normal service on this PLMN.

7.1.3. ARCHITECTURE

Selecting a PLMN and a cell is but one side of the management of mobility. The goal of the network as far as location management is concerned is to prepare for the routing of calls toward the subscribers, taking account of their movements. To that end, the network must memorise for each subscriber (very precisely for each SIM) whether he is known to be in some place or not (he is said to be **registered**), and if so, in which location area. This information is retrieved when a call toward the user must be set up, as will be explained in Chapter 8. Because a location area is mandatorily in GSM wholly included in a single MSC area, the stored information is sufficient for routing the call up to the MSC which will be in charge of the communication. Furthermore, the knowledge of the precise location area (there may be several location areas within an MSC area) allows to restrict the paging to the corresponding cells.

A simple solution to the basic location management issue could consist in storing in a database the identity of each subscriber together with an indication on whether or not he is registered, and if so, where to find him.

Indeed, the canonical architecture of GSM identifies such a database, the Home Location Register (HLR). This function is separated from the routing function itself, which consists in choosing and reserving circuits to obtain a continuous connection between users that desire to be in communication. In GSM, the main actor on the mobile user side for routing and communication management functions is the Mobile services Switching Centre (the MSC). But the canonical architecture is a bit more complex, and before describing it, it is interesting to look at to the reasons why.

Every telecommunication system includes a database containing a variety of information concerning each subscriber, such as the subscription limitations, the services subscribed for, the states of the supplementary service activation, or the information needed for the management of the charging information. In GSM, the same information exists, plus some which is specific such as the information related to the confidentiality functions.

In a fixed network, each subscriber is connected to one local switch, for a long time. Every call involving this subscriber, whether an originating or a terminating call, goes through this switch. This is then the natural place for the storage of the subscriber related information. In a system dealing with moving subscribers, there is no such natural place for the storage of subscriber parameters. However, the two kinds of data to

be stored (location information and subscriber data) call for a common storage solution. This is the choice made in GSM, and the HLR is the database for both sets of information.

If location information is needed only for the establishment of mobile terminating calls, the rest of the information is needed at various moments during any call. Basically, it is the visited MSC, the one in charge of a mobile subscriber engaged in a call, which needs these pieces of information. Then an important signalling load would result if the MSC had to interrogate the HLR each time it needs some piece of information. To avoid this signalling load, the data record of a subscriber is copied in a database close to the MSC while this subscriber is registered in a location area controlled by the MSC. The database, the VLR (Visitor Location Register), will be ignored for the moment as an entity separate from the MSC: we will speak of an MSC/VLR. The distinction will be dealt with later, and our approach explained.

This capacity of temporary storage in the MSC/VLR allows some distribution of the function of location information storage. The HLR needs only store information concerning the MSC/VLR in which area where the subscriber currently is. The identity of the precise location area is stored in the MSC/VLR, together with a copy of the remaining subscriber related information. This is sufficient to get the routing information needed to deal with an incoming call, and this is all the HLR needs.

This temporary storage in the MSC/VLR introduces new functions. The subscriber information has to be copied when the subscriber enters a new MSC/VLR area. Conversely, the corresponding record has to be erased in the previous MSC/VLR in which the subscriber was registered. Some mechanisms are needed for maintaining the consistency between what is stored in the HLR and what is stored in the MSC/VLRs, including the case of a failure resulting in a loss of stored information.

7.1.3.1. Functions

As defined in the *Specifications*, the HLR is basically an intelligent database used to store the location information and the subscriber related information needed for providing the telecommunication services. The HLR has no switching capability. It is connected to the other entities of the network and switching subsystem (NSS) through signalling means, as discussed in Chapter 2. The HLR is not a simple database which can accept only "store" or "retrieve" orders. In fact, the HLR completely manages the

location information in the network: for instance, it has to tell the old MSC/VLR to erase a subscriber record when this subscriber is registered under a new MSC/VLR.

Functionally, we could say that there is a unique HLR function system-wise, possibly distributed on several equipments. In practice, be it only for operation reasons, one HLR function is implemented in each PLMN. There again, the HLR function for a PLMN can be implemented in a single equipment or distributed among several equipments. Both approaches are allowed, and used. Note that the usage of the term HLR may refer to the function, possibly encompassing several equipments, or to a single equipment. In most of the cases, this ambiguity causes no understanding problem.

The HLR has many different roles. What concerns us in this section is the management of the subscriber mobility. Functions of the HLR related to, e.g., the management of the confidentiality data, or the management of the supplementary services shall be described respectively further on in this chapter, and in Chapter 8.

The Visitor Location Register

While we have introduced the term VLR, the corresponding concept has been somewhat masked. Most readers will have also noted that the VLR was not even allocated an icon. We have to answer for this rather off-hand treatment.

In the canonical GSM architecture, what has been referred up to now as the MSC/VLR consists of two disjoint functional entities, the MSC itself and a database, the VLR. The MSC is defined as the switching function in charge of the management of the calls, and the VLR as the database where subscriber information are temporarily stored for those subscribers which are registered under a MSC connected to the VLR.

A VLR can manage the subscriber data for one or several MSCs, and can be an equipment physically distinct from a MSC. The reason why the two functions are split is not so much because of this possible implementation in distinct equipments, but because of the option to have a VLR for more than one MSC. The point to analyse is then why should such a choice be taken.

From an architectural point of view, the VLR can be seen from three different vantage points:

- a first approach is to consider the set of the HLR and VLR as a single distributed database. The distinction between the HLR

parts and the VLR parts becomes a matter of internal architecture of this database. This would correspond to an approach where the VLR is introduced solely for signalling load distribution, and a VLR would naturally be serving several MSCs;

- the opposite point of view is to consider the VLR as fulfilling a set of ancillary tasks to the MSC, including the management of the visited subscriber database and the corresponding dialogues with the HLR. It is then naturally a part of the MSC, and there are little reason for having a VLR connected to several MSCs;
- a third approach is to consider the VLR as a truly independent entity, having tasks of its own, with added value compared to the natural roles of the HLR and the MSC.

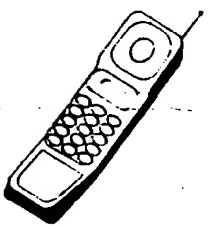
The exact philosophy of the functional split between the VLR and the HLR on one side, and between the MSC and the VLR on the other, can be determined by looking in details of the corresponding protocols. In so doing, it becomes apparent that the cut between the MSC/VLR and the HLR can be considered as a minimum, whereas the cut between the VLR/HLR and the MSC seems not to follow strongly directive lines. If the VLR/HLR protocol can be easily presented as an MSC/HLR protocol, it is obviously impossible to do so with the MSC/VLR protocol (the proof of the first statement will be found in this book, whereas it would be too long to justify the second; the interested reader can look to the *Specifications* to make up his mind). This militates strongly for the second approach, that we have followed in this book. It is often invoked as a counter argument that the split between the VLR and the MSC is related to the Intelligent Network (IN) approach for building switch equipments. When looked at closer, though, it is obvious that the two philosophies are somewhat different: in an IN approach, the MSC would have no high level function and the VLR would be where all the complex protocols are dealt with. This is very different in the *Specifications*: the MSC deals indeed with most of the complex protocols. The VLR is neither a pure database (in which case, the VLR-MSC interface would be very close to an interface between a MSC and a HLR if temporary storage was not used), nor the true call manager controlling a rather dumb switch (this would be the IN approach). Implementations of a GSM network based on an IN approach are nevertheless possible and implemented, but the split between the "dumb" Switching Service Point and the "intelligent" Service Control Point is not based along a MSC-VLR line. While the architectural split as described in the *Specifications* is for these reasons certainly not the last say in the domain, GSM is considered as one of the first "intelligent" networks and concepts such as

the interrogation of a centralised HLR are the first bricks for the construction of a full-blown intelligent network structure.

This situation is the main reason why the authors of this book decided not to describe the VLR as an entity separated from the MSC. This position is supported by the fact that up to now *all* the switch manufacturers have chosen to develop a combined MSC/VLR, and none offer the possibility to physically split them up.

Let us see rapidly some of the points of the *Specifications* among the casualties caused by our approach. If separated, MSC and VLR are connected through the SS7 signalling network, and the signalling procedures for the corresponding dialogue are specified in the MAP (they constitute the MAP/B protocol). The MAP/B protocol represents a major part of the MAP specification, at least as far as the number of pages is concerned. Because of the reasons explained above, it will not be treated here.

The Mobile Station



The mobile station holds a starring role in location management, for obvious reasons. It is at the origin of the location information, and in fact deals with problems posed by mobility almost entirely on its own in idle mode. An important point which may have been less evident up to now is the respective role of the SIM and of the rest of the mobile station, known in the *Specifications* as the mobile equipment.

The assumption in the *Specifications* is that the mobile equipment does not hold in a non-volatile memory any information specific to its user. Still, mobile station manufacturers are not prevented from doing so, but the system can work properly without such a memory. The converse point is that the SIM holds this information, and in particular many related to the mobility management, whether for handling location or security related information.

The mobile equipment contains however some information of some temporal scope, such as the list of forbidden location areas for national roaming, or lists of beacon frequencies for different PLMNs. This information is lost when the mobile equipment is switched off. The choice of what is in the SIM was a compromise between memory consumption (a scarce resource for a SIM) and keeping as much as possible potentially useful information over a switched-off period.

The SIM

The SIM has already been presented in the very first chapters. What is of interest here is the information it contains in relationship with location management. In this area, the SIM is but a passive information container, so listing the relevant fields will give us all the functional description we need:

- The update status;
- A location area identity;

These fields usually contain the result of the last location updating attempt, and the location area where it was done; their main purpose is to avoid a location updating attempt in some cases when the cell selected after switch-on is in the same location area.

- A list of beacon frequencies ("BCCH information");

We have already met this list. The *Specifications* are not clear about which PLMN it relates to. In order for this list to achieve its aim; i.e., to speed up the initialisation time after switch-on, it should pertain to the home PLMN. Another interpretation is that it should be the last list received from the serving network.

- The *forbidden* PLMNs list ("forbidden PLMNs");

The function of this list was described in the section of this chapter relative to PLMN selection. It should be noted that it is an *ordered* list, with entries sorted by order of introduction, because of the requirement to replace the oldest entry when the list is full.

- The *preferred* PLMN list.

The function of this list was also described when dealing with PLMN selection.

7.1.3.2. Protocols

The protocols belonging functionally to the Mobility Management plane are the one held between the HLR and the MSC/VLRs, the MSC/VLR and the mobile station, and between the mobile equipment and the SIM. To allow full roaming, it is of utmost importance that every HLR be able to exchange information with every MSC/VLR throughout all the PLMNs of the system (and possibly with switches from other types of networks if SIM-roaming is implemented). A HLR must also be able to dialogue with all the entities that want to get information about

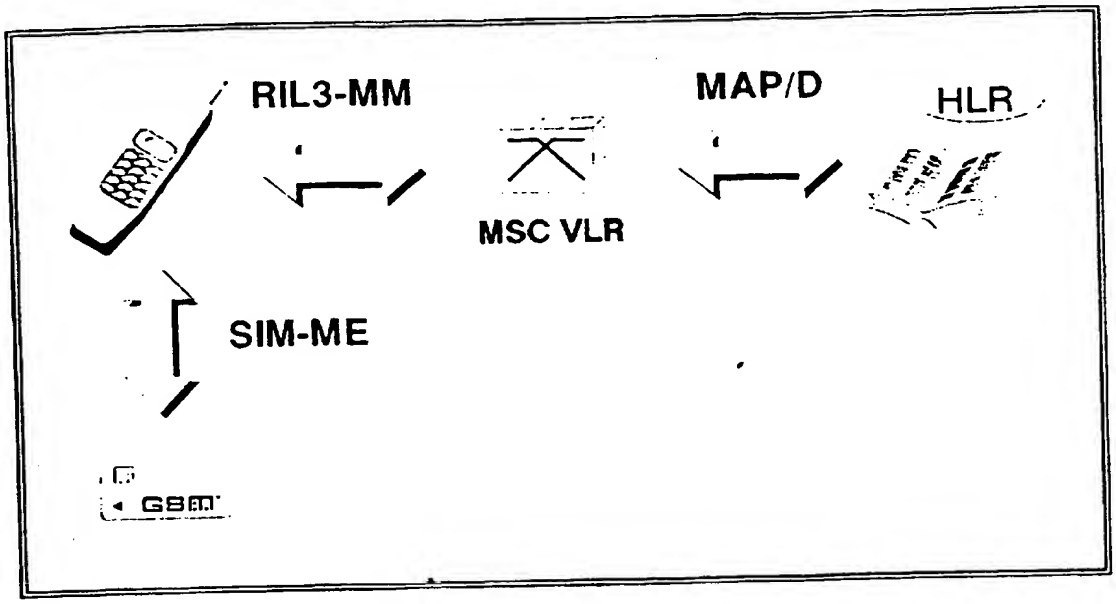


Figure 7.4 – Location Management Protocols

MM protocols involve the location registers (VLR and HLR), which communicate between themselves and with the mobile station, in which the SIM plays a prominent part.

the subscribers, mainly location information with the aim of setting up calls toward these subscribers. This will be dealt with in Chapter 8.

The protocol between the HLR and the MSC/VLRs is supported through the world-wide signalling network, the signalling system n°7 (SS7), as described in Chapter 5. The application protocol for the dialogues between a HLR and a MSC/VLR is part of the MAP (Mobile Application Part). In this book, we will call it the **MAP/D** protocol.

The MS-MSC protocol is called the **RIL3-MM** protocol (for Radio Interface Layer 3, Mobility Management). It uses the MS-MSC signalling connection provided by the RR layer, as seen in Chapters 5 and 6.

The **SIM-ME** protocol is limited to read and write commands as far as we are concerned here, and corresponding messages will not be cited in the text.

Figure 7.4 summarises the simple architecture of the protocols needed for location management.

7.1.4. THE LOCATION UPDATING PROCEDURES

The main procedure of interest for location management is the location updating procedure, which is triggered by a mobile station to

update the location data of its subscriber. For various reasons, slightly modified versions of this procedure are used for different related purposes. These variations are also described in this section.

7.1.4.1. The Basic Procedures

The location information is stored in two different places in the GSM infrastructure, the HLR and the visited MSC/VLR. In fact the same information is known in three different places in the system, the mobile station (and more explicitly the SIM) being the third place. This information may change, and various procedures are needed to keep the consistency between the three entities.

The normal reason for a change is when the mobile station decides that the location area best fit to serve its subscriber must be changed. Then the mobile station notifies the MSC/VLR to which the new cell belongs. This MSC/VLR may be the same as before, if it controls both the previous and the new location area, or a new MSC/VLR. In the latter case, the MSC/VLR notifies in turn the HLR, which notifies the previous MSC/VLR. There are other cases where an inconsistency may appear, for instance when stored information is lost in the MSC/VLR or the HLR as a result of some hardware or software failure. Then procedures may be run to correct the failed database using information in other equipments.

In order to cover all theses cases, the following elementary procedures have been specified (see figure 7.5):

- Updating of the MSC/VLR storage at the request of the mobile station;
- Updating of the HLR storage at the request of the MSC/VLR;
- Cancellation of a subscriber record in a MSC/VLR at the request of the HLR;

The Mobile Station to MSC Location Updating Procedure

This procedure is part of the RIL3-MM protocol. It requires a radio connection, as for any dialogue between the mobile station and the network. The establishment of such a connection is a function of the Radio Resource Management functions, described in Chapter 6. This establishment has almost nothing specific to the location updating procedure.

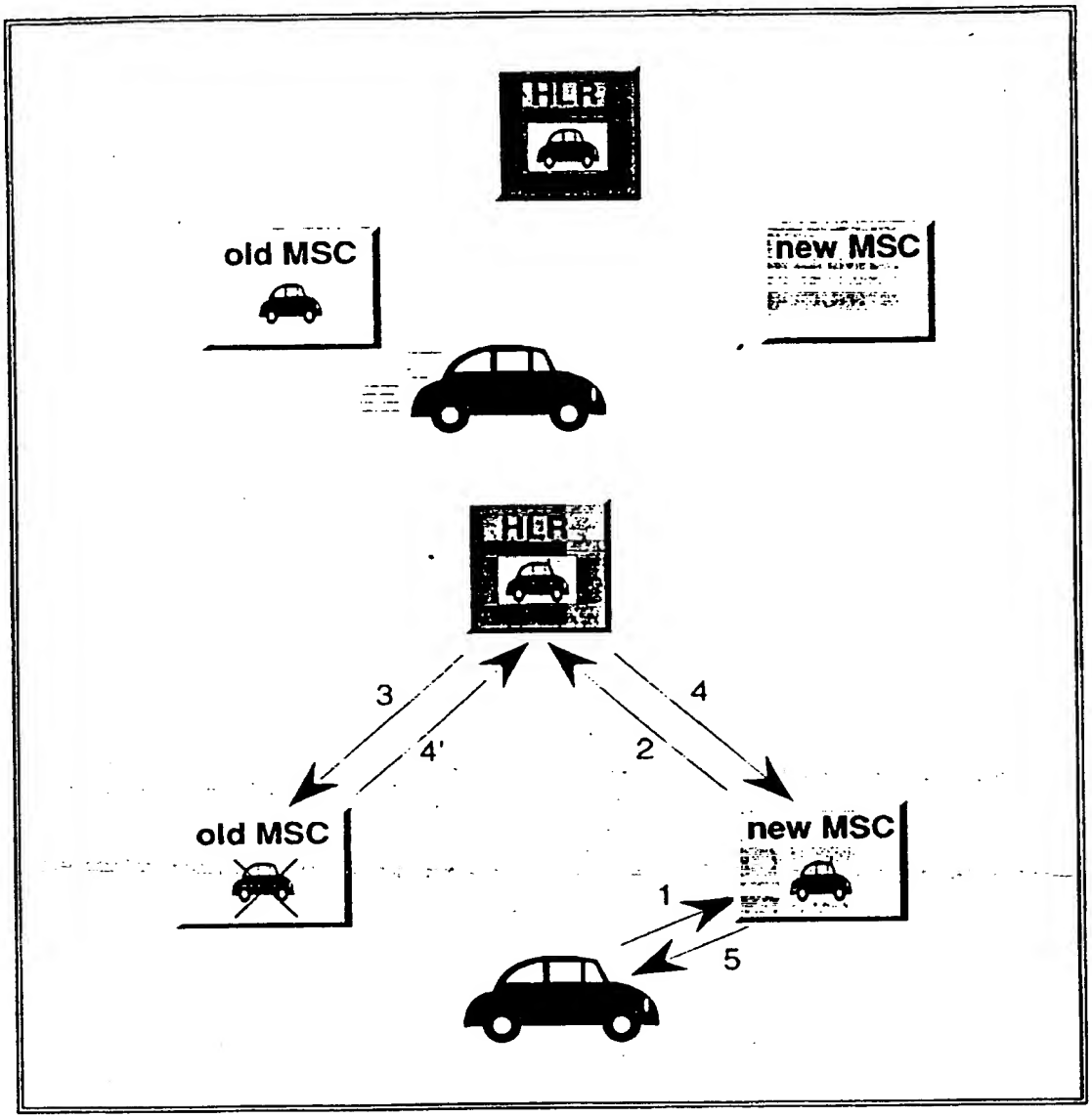


Figure 7.5 – Main elementary location updating procedures

In order to change from the old situation (shown on top) to the new situation (bottom),
the mobile station takes the initiative of location updating (1),
but the HLR, upon subsequent request from the new MSC/VLR (2),
takes care of cancelling the old record in the previous MSC/VLR (3, 4')
in parallel with confirming the updating in the new MSC/VLR (4),
which in turn acknowledges the mobile station request (5).

The MS-MSC location updating procedure is basically very simple, and consists of a request (a location updating request) and an answer. The request is carried by the RIL3-MM LOCATION UPDATING REQUEST message. This message contains mainly the information necessary to identify the subscriber.

The MSC/VLR may answer autonomously in some cases, or alternatively may have to update the HLR first, with the procedure described in the next section. There is one case when the MSC/VLR cannot do otherwise than answer on its own: when it cannot reach the HLR for lack of any roaming agreement between the two operators. This case is not addressed in the *Specifications*, though it can happen. The answer of the MSC/VLR is necessarily negative and must be chosen to ensure that the mobile station will search other PLMNs (for instance, by sending the cause "PLMN not allowed").

The usual "normal case" when the MSC/VLR can answer on its own is when the subscriber is already registered in the database of the MSC/VLR. The response in that case is usually positive, and can be negative only in case of national roaming restriction (cause "location area not allowed for national roaming", used only in DCS1800); regional subscription cannot lead to a negative MSC/VLR answer without HLR involvement, since the MAP restricts regional subscription to be offered on a per-VLR basis.

National roaming then merits some attention. The rule is that if the mobile station belongs to an unwanted PLMN, and if the requested location area is restricted, the MSC/VLR is entitled to directly answer negatively. Most of the time the MSC/VLR will not be able to contact the HLR, and this would be a particular case of the situation mentioned above, with the difference that the cause sent to the mobile station is specified.

When the MSC/VLR needs to contact the HLR of the subscriber, it must first know which HLR is concerned. Subscribers are identified for the internal business of GSM by a number, the IMSI (International Mobile Subscriber Identity). This number is provided by the mobile station anytime it accesses the network (the number is not always given directly, see the notion of TMSI, page 484). The IMSI is so specified that the MSC/VLR is able to derive the identity of the subscriber's home PLMN, and possibly more information on the HLR equipment in charge of the subscriber. Figure 7.6 shows the IMSI structure. With the help of the relevant translation tables, the MSC/VLR is then able to derive the SS7 address to which the location updating request must be sent. In practice, the HLR can usually be identified by looking at the most significant digits of the IMSI following the mobile country code and mobile network code. However, this possibility is usually only used inside the home PLMN country. PLMNs of other countries route their messages using the IMSI as a global title, towards a gateway entity in the home PLMN country. There the global title can be translated in the Signalling Point Code of the right equipment in the right PLMN, as explained in Chapter 5.

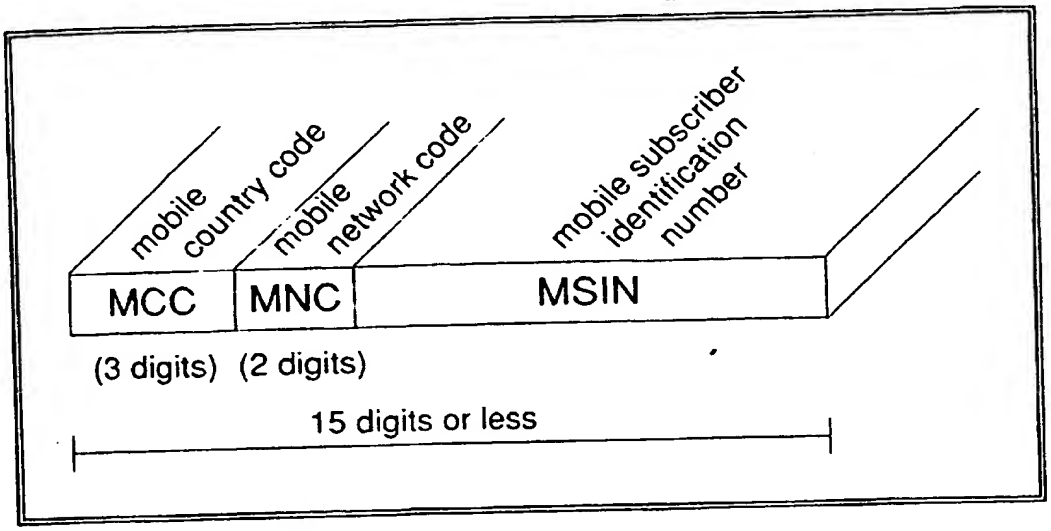


Figure 7.6 – The structure of an IMSI

An international Mobile Subscriber Identity consists of three parts:
the Mobile Country Code (MCC), identifying a country;
the Mobile Network Code (MNC), identifying a PLMN within this country; and
the Mobile Subscriber Identification Number, identifying a subscriber
within this PLMN, using no more than 10 digits.

Before giving the answer to the mobile station, in an RIL3-MM LOCATION UPDATING ACCEPT message or an RIL3-MM LOCATION UPDATING REJECT message, the MSC/VLR may proceed to some actions like authentication or ciphering setting. These procedures are detailed in the second part of this chapter.

The answer provided by the MSC/VLR, possibly after contact with the HLR, is either the RIL3-MM LOCATION UPDATING ACCEPT, thus indicating that the subscriber is effectively registered in the new location area as required, or RIL3-MM LOCATION UPDATING REJECT, with a suitable cause, such as "PLMN not allowed" or "location area not allowed". The first indicates that the subscriber has no subscription entitlement for service in the visited PLMN, whereas the second, which can be used only in the home PLMN, indicates the subscriber has no subscription entitlement for service in the location area. The HLR does not indicate which cause to use (the MAP provides only one cause "roaming not allowed"). Though not indicated in the *Specifications*, the cause in the RIL3-MM message should be "location area not allowed" if the visited MSC/VLR belongs to the home PLMN, and "PLMN not allowed" otherwise. The mobile station reacts on the receipt of a negative answer as explained in the section on cell and PLMN selection.

The MSC/VLR to HLR Location Updating Procedure

This procedure is run when a mobile station asks for registration under a new MSC/VLR. It can also be run when the HLR has suffered a failure, and request the MSC/VLR for a confirmation of the subscriber location as soon as the mobile station is in contact.

The request is conveyed in a MAP/D UPDATE LOCATION message. This message carries (among other information) the subscriber identity and enough information for the HLR to know how to find routing data for the setting up of a mobile terminating call, i.e., the SS7 address of the MSC/VLR. It does not convey the identity of the precise location area. As an option supported by the standard, the routing data can be included at this point, but we will see in Chapter 8 that this is not what is usually done.

The HLR determines whether to accept to register the mobile station in the new MSC/VLR or not by consideration of the subscription limitations of the user. If the subscriber is entitled to normal service in the requested VLR area, the answer is positive, the HLR updates its memory and triggers a location cancellation procedure with the previous MSC/VLR (if the mobile station was actually registered, and in another MSC/VLR). If the subscriber cannot be granted normal service, the HLR deregisters the subscriber: it updates its memory to indicate that the location of the mobile station is unknown, and it triggers a location cancellation procedure if applicable. The answer from the HLR is carried to the MSC/VLR in a MAP/D UPDATE LOCATION RESULT message.

If the requesting MSC/VLR receives a negative answer, it erases all information relative to the subscriber. Otherwise and if necessary, it enters the subscriber in its database. Normally, the HLR will then provide the subscriber information the MSC/VLR needs. This is done by sending a MAP/D INSERT SUBSCRIBER DATA message, which is acknowledged by the MSC/VLR through a MAP/D INSERT SUBSCRIBER DATA RESULT message. The same procedure may also be used when some change in the subscriber's information stored in the HLR has occurred, for instance at the request of the subscriber. Alternatively, depending on the nature of the modification, a procedure composed of the MAP/D DELETE SUBSCRIBER DATA and the MAP/D DELETE SUBSCRIBER DATA RESULT messages may be used.

The HLR to MSC/VLR Location Cancellation Procedure

Location cancellation from HLR to MSC/VLR is a very simple procedure, composed of a request carried by a MAP/D CANCEL LOCATION

message and the corresponding acknowledgement carried in a MAP/D CANCEL LOCATION RESULT message. The HLR does not wait for this acknowledgement to confirm location updating to the other MSC/VLR who triggered the change of location.

The MSC/VLR to HLR Deregistration Procedure

A MAP/D procedure opens the possibility for the MSC/VLR to ask the HLR for the deregistration of a given subscriber. There is no identified case in GSM for this procedure, and it is indicated as not used in a laconic sentence in the first page of TS GSM 09.02.

This procedure is the simple order/acknowledge procedure. The request is carried by a MAP/D DEREGISTER MOBILE SUBSCRIBER message, and acknowledged by a MAP/D DEREGISTER MOBILE SUBSCRIBER RESULT message.

7.1.4.2. Periodic Location Updating, and Database Failure Recovery

An HLR, or an MSC/VLR, may suffer a failure such that a part of its database is damaged. These equipments are usually implemented in a secure way with back-up systems, allowing the database to be restored to some consistent state. However, there are cases when the restored database is no longer up-to-date, e.g., because updates having occurred between the last back-up and the failure have been lost.

Since such conditions may affect the possibility to set up calls correctly, a number of mechanisms are described in the *Specifications* to improve the situation. The general mechanism is as follows: in a first step, the affected equipment will mark the uncertain information as such. Then it will warn other network entities (those with which it knows to be sharing information) about the unclear status of its information contents. Consequently, these entities will also mark the corresponding records for checking. To avoid overloading the signalling system, no attempt is made to restore the consistency right away: the recovering equipment would probably not survive a high peak in signalling load! A given subscriber record (say Christian's) is corrected only when some event happens concerning this subscriber, such as a radio contact initiated by the mobile station, or a mobile terminating call attempt. If after some time nothing has happened, Christian is deregistered.

Obviously, the best place to stay informed about the subscriber position is the mobile station, and a radio contact is at the end of the day

the only safe way to restore consistency. Periodic location updating has thus been introduced to ensure regular radio contact at the initiative of the mobile station.

Basically, periodic location updating refers to the requirement that the mobile station contacts regularly the network when in normal service. This is done automatically by the mobile station, and takes the form of a location updating procedure. The period is under control of the network which broadcasts its value. The possibilities range from 6 minutes (in which case the network is probably completely overwhelmed by location updating procedures!) to slightly more than 24 hours. In addition, the infinite value is included: in other words, the network can suppress completely periodic location updating if so wished. The choice of periodicity is the operator's, and is typically a trade-off between the quickness of recovery after a failure (therefore the time when it may be impossible to serve mobile terminating calls for some subscribers) and the traffic load due to periodic location updating. This clearly depends a lot on the reliability of the MSC/VLR and HLR: if the mean time between failures is very long, there is no point in incurring the load and cost of periodic updating (except if implicit detach is also used, see page 476). On the other hand, the periodicity can be increased by the operator after a failure.

Let us now see in more details how the location registers cope with losses of information after failure, and how in general location data is recovered.

MSC/VLR Failure

When a MSC/VLR suffers a database failure, it first restores its state to some previously saved state, and after this recovery marks all its records as to be checked with the mobile station and with the HLR. Then it sends a MAP/D RESET message to all HLRs for which it still has a subscriber in its tables. The MSC/VLR will eventually notice that a mobile station is missing in its records when some service must be provided to the subscriber, and at the latest when the mobile station performs a periodic location updating. In the reverse situation (i.e., a mobile station still recorded when it should not be), the correction may take longer.

In the case of a request from the mobile station (e.g., a call set-up, but not a location updating of some sort), the MSC/VLR will notice for instance that the corresponding subscriber is not in its table though the mobile station thinks so. The MSC/VLR reacts then by requesting the mobile station to perform a location updating. This is done with a

rejection with cause "IMSI unknown in VLR". Because this will also trigger a location updating from MSC/VLR to HLR, consistency will be reached anew. A milder case is when the mobile station calls from a location area which is different from the one in which the MSC/VLR thinks it is registered. In this case, the MSC/VLR simply corrects its record.

In the case of a mobile terminating call, the MSC/VLR may notice a problem if it receives a request from the HLR concerning a mobile station which is not in its table though the HLR obviously thinks so. This may happen either because of an MSC/VLR failure or because of an HLR failure. In all cases, the MSC/VLR enters the subscriber in its tables, and asks the HLR for the subscriber information via a MAP/D SEND PARAMETERS message, which is answered with a MAP/D SEND PARAMETERS RESULT message from the HLR. The location area information is in this case still missing. Henceforth, and until actual contact with the mobile station, the paging is done in all cells of all location areas controlled by this MSC/VLR. Note that if the error was in the HLR, that is to say if the mobile station if effectively not under the MSC/VLR as the HLR knows, the information in the network is inconsistent: after a mobile terminating call attempt, the subscriber is registered in two MSC/VLRs (MSC/VLR–1 where it is really, and MSC/VLR-2 where the HLR imagines it is), and the HLR stores the identity of the wrong one. The error will be corrected in the HLR and in MSC/VLR-2 when a location updating procedure is performed by the mobile station. However, MSC/VLR-1 will remain in error by keeping the subscriber in its database. The cleaning-up can only be done by an internal mechanism in the MSC/VLR, erasing all records of subscribers which have been inactive for more than some long time, e.g., a month.

HLR Failure

Restoration of the HLR is a bit more complex, because the HLR is not necessarily contacted in the case of a periodic location updating or a mobile originating call. To enforce this contact, the HLR sends a MAP/D RESET message to all MSC/VLRs in which at least one of the HLR subscribers is known to be located, as indicated by the salvaged records. The MSC/VLRs will mark all corresponding records as to be checked with the HLR: the next radio contact will then trigger a location updating procedure from MSC/VLR to HLR, thereby correcting the HLR records.

In the case of a mobile terminating call, as we have already seen two paragraphs above, the interrogated visited MSC/VLR aligns its database with the (recovered and unchecked) HLR data. At least, this is

all that the MAP/D protocol, as specified, enables it to do. However, HLR restoration methods described in TS GSM 03.07 include a different mechanism, by which the HLR would indicate to the MSC/VLR that the interrogation concerns a mobile station for which data is unsure: the MSC/VLR would in that case know whether it is better to take the HLR as a reference or to rely on its own state and answer accordingly. However, this mechanism has not been implemented in the phase 1 MAP protocols.

7.1.4.3. The IMSI Attach and Detach Procedures

The location updating procedure has two very close siblings: the periodic location updating procedure, which has just been described, and the *IMSI attach* procedure: On the radio path, both procedures are almost identical to a location updating procedure. They differ from it almost only by the events that trigger them. These events are such that these "location updating" procedures usually appear as a request to be registered in the location area where the subscriber is *already* registered. Hence in most of the cases, the HLR is not concerned with these procedures.

The strange names of the *IMSI attach* and *IMSI detach* procedures are due to the accidents of the standardisation, and are of little use to understand their meaning. The best way to understand these procedures is to explain what purpose they satisfy.

When a mobile station is switched off (or when the SIM is removed by the user), the calls toward the corresponding subscriber can no more be completed. Important resources are then consumed for nothing: as will be seen in Chapter 8, a circuit is established between the caller and the MSC in charge of the called mobile station, and the paging procedure is performed, all to no avail. Worse (from one point of view), the established circuit is not paid for. As will be seen when dealing with the routing of calls, the establishment of the first part of the circuit (before HLR interrogation) cannot be avoided. It is a different story with the second portion, between the point where HLR interrogation is done and the visited MSC.

To alleviate this useless load (and cost), the *IMSI detach* and *IMSI attach* mechanism has been introduced. Basically, the subscriber's record in the MSC/VLR contains a binary information indicating whether it is useful or not to try to complete a call toward this subscriber. This information makes it at least possible to economise on paging. It may also prevent the establishment of a part of the call. The *IMSI detach* procedure will set this bit to "not useful to try", whereas the *IMSI attach* procedure will do the reverse. The mobile station triggers an *IMSI detach* when it

goes inactive, and either a location updating procedure (if in a new location area) or an *IMSI attach* procedure when it comes back (in the same location area).

The MAP specifications include two different ways for the management of this feature in the infrastructure: either the information is only stored in the MSC/VLR (the mobile station staying registered in this MSC/VLR as far as the HLR is concerned), or the subscriber is simply deregistered in the HLR and its record cancelled in the MSC/VLR. In fact, only the first option is allowed for GSM, since the deregistration procedure is not used.

With this first option, paging can obviously be prevented. The second part of the circuit establishment may also be prevented, but not so obviously. Moreover, even when possible, it is an option to prevent it or not. As will be seen in Chapter 8, the basic scenario of a mobile terminating call set-up attempt requires an interrogation of the visited MSC/VLR by the HLR before the latter provides the information necessary for the continuation of the routing. This phase allows the visited MSC/VLR to reject the call on the basis of the attach status before the costly set up of the traffic circuit. If it does so, call forwarding if applied can potentially be controlled by the HLR. Another possibility is that the visited MSC/VLR accepts the call, and applies the call forwarding itself if required.

To complete the option list, the support of the attach/detach feature is a network option. It is allowed that a PLMN, or a portion thereof, does not provide this facility. This is indicated to the mobile stations on a cell basis in the broadcast information. The choice between all these options and sub-options lies with the visited MSC/VLR, since the HLR does not intervene. If attach is indicated as supported in the current cell, the *IMSI detach* procedure is used by the mobile station to indicate that it (more exactly the SIM) will go inactive.

The *IMSI detach* procedure is an example of a procedure reduced to its bare bones: it consists of a single message from the mobile station to the visited MSC/VLR, the RIL3-MM IMSI DETACH message. This message is not acknowledged, simply because it has been considered that the mobile station is typically switched off, or more generally not in a position to receive an answer from the network. The *IMSI detach* procedure must use a radio connection, as any RIL3-MM procedure. This connection is either established for the purpose of the detach, or may pre-exist. The connection can be abandoned by the mobile station immediately after the sending of the RIL3-MM IMSI DETACH message. The mobile station keeps no track of having asked for a detach (for instance by storage in the SIM): the state of the attach/detach information in the network is not monitored by the mobile station.

If attach is indicated as supported in the cell the mobile station has chosen at switch-on (or SIM insertion), and if the mobile station knows the subscriber is already registered in the same location area, it starts an IMSI attach procedure, that is to say (except for a negligible detail) a location updating procedure. It should be noted that the attach procedure may happen even if there was no request for detach beforehand (because the network did not require it at the time, or simply because it was not physically possible, e.g., in case of a loss of coverage before switch-off). This is consistent with the specification that the mobile station does not monitor the attach/detach status.

These procedures would then have been better called "Subscriber deactivation" and "Subscriber reactivation". From a very slightly different point of view, these procedures are very close functionally to the call forwarding supplementary services in the case where the mobile station is not deregistered. The *IMSI detach* procedure can indeed be understood as an automatic activation of an unconditional call forwarding (toward an announcement, or a specific number), and the IMSI attach as the corresponding deactivation procedure. It can therefore be said that, from the point of view of the purpose of these procedures, of their *why*, they have little to do with the location updating, but are akin to call forwarding. From the point of view of the mechanisms they use, that is to say from their *how*, these procedures are close siblings to the location updating procedure.

7.1.4.4. Automatic Detach by the Network

As explained above, the attach status can be modified by the mobile station, through the *IMSI attach* and *IMSI detach* procedures. A past debate in the GSM committees was whether an additional mechanism may be added, consisting in putting a subscriber in the detach state, or deregistering him, if nothing has been heard of him for more than a given period. The debate ended more or less against this feature. It can be nevertheless used, since it requires only existing signalling procedures and otherwise affects only the MSC/VLR.

Thanks to periodic updating, the mobile station of a subscriber when switched-on must indeed contact the network at least once every so often. Thanks also to the non-storage of the detach status in the SIM, the mobile station will always make itself known when re-activated: there is no period after switch-on during which the mobile station cannot be called because of a forceful detach by the network.

The advantage of a network time-controlled detach is that the detaching may happen even in the cases where the mobile station is physically unable to send an RIL3-MM IMSI DETACH message.

This feature is not mentioned anywhere in the *Specifications*. However it would work without side effects, and it is known that a number of operators intend to use it.

7.2. SECURITY MANAGEMENT

Radio transmission is by nature more prone to eavesdropping and fraud than fixed wire transmission. Listening to communications is easy and does not require access to special locations. Impersonating a registered user (and therefore having him foot the bill!) can also be very easy if specific protection means are not provided. Analogue systems have indeed suffered from such problems during the 80's. GSM had to bring significant improvements in these matters.

7.2.1. THE NEEDS

The security-related functions of GSM aim at two goals: first, protecting the network against unauthorised access (and at the same time protecting the users from fraudulent impersonations); second, protecting the privacy of the users.

Preventing unauthorised accesses is achieved by means of authentication, i.e., by a secure check that the subscriber identity provided by the mobile station corresponds to the inserted SIM. From the point of view of the operator, this function is of paramount importance, in particular in conjunction with international roaming, where the visited network does not control the subscriber's record... and his ability to pay.

Preserving the privacy of the users is achieved through different means. Transmission can be ciphered to prevent eavesdropping of communications on the radio path. Most of the signalling can also be protected in the same way, preventing third parties from knowing who is being called, for instance. Finally, the replacement of the subscriber's identity by a temporary alias is another mechanism to convince third parties that listening on the radio path is useless for tracing GSM subscribers. Since most of the calls involving a GSM user go through the fixed network, the designers of GSM did not aim at a level of security

much higher than that of the fixed trunk network. Mechanisms to ensure privacy have only been introduced for the radio path. Within the infrastructure, communications are transmitted in clear text, as they are in the PSTN.

It is important to note at this stage that all the security mechanisms of GSM are under sole control of the operators: the users have no possibility to affect whether authentication, encryption, etc. are applied or not. Moreover, users are not necessarily aware of what security features are used. Conversely, these security services are not usually subscribed-for. The *Specifications* leave a lot of flexibility to apply them in various conditions. Some harmonisation is however desirable, and is settled for the GSM900 operators for instance by discussions within the GSM MoU.

7.2.2. THE FUNCTIONS

7.2.2.1. Authentication

A simple authentication method is the use of a password (or a PIN code—Personal Identity Number). The level of protection achieved by such a method is very low in a radio environment, since listening once to this personal code is enough to break the protection. GSM does make use of a PIN code in conjunction with the SIM; this PIN code is checked locally by the SIM itself, without transmission on the radio interface. But in addition, GSM uses a more sophisticated method, consisting in a layman's words in asking a question that only the right subscriber equipment (in that case, the SIM) may answer. The crux in this method is that a huge number of such questions exist, and that it is therefore very unlikely that the same question would be used twice.

More precisely, the question takes the guise of a number, called *RAND* in the *Specifications*, whose value is drawn randomly between 0 and $2^{128}-1$ (something like a few millions of milliards of milliards of milliards of milliards!). The answer, called *SRES* in the *Specifications*—i.e., Signed RESult in cryptographic terminology—is obtained as the outcome of a computation involving a secret parameter specific to the user, and called *Ki* in the *Specifications* (see figure 7.7). The secrecy of *Ki* is the cornerstone on which all the security mechanisms are based. We will see that it is stored in a very protected way; for instance a subscriber cannot know his *Ki*. The algorithm describing the computation is referred to as algorithm A3 in the *Specifications*, but its specification cannot be found there. In fact, the design choices of GSM, both in the mobile station and in the infrastructure, allow A3 to be operator-dependent while allowing full

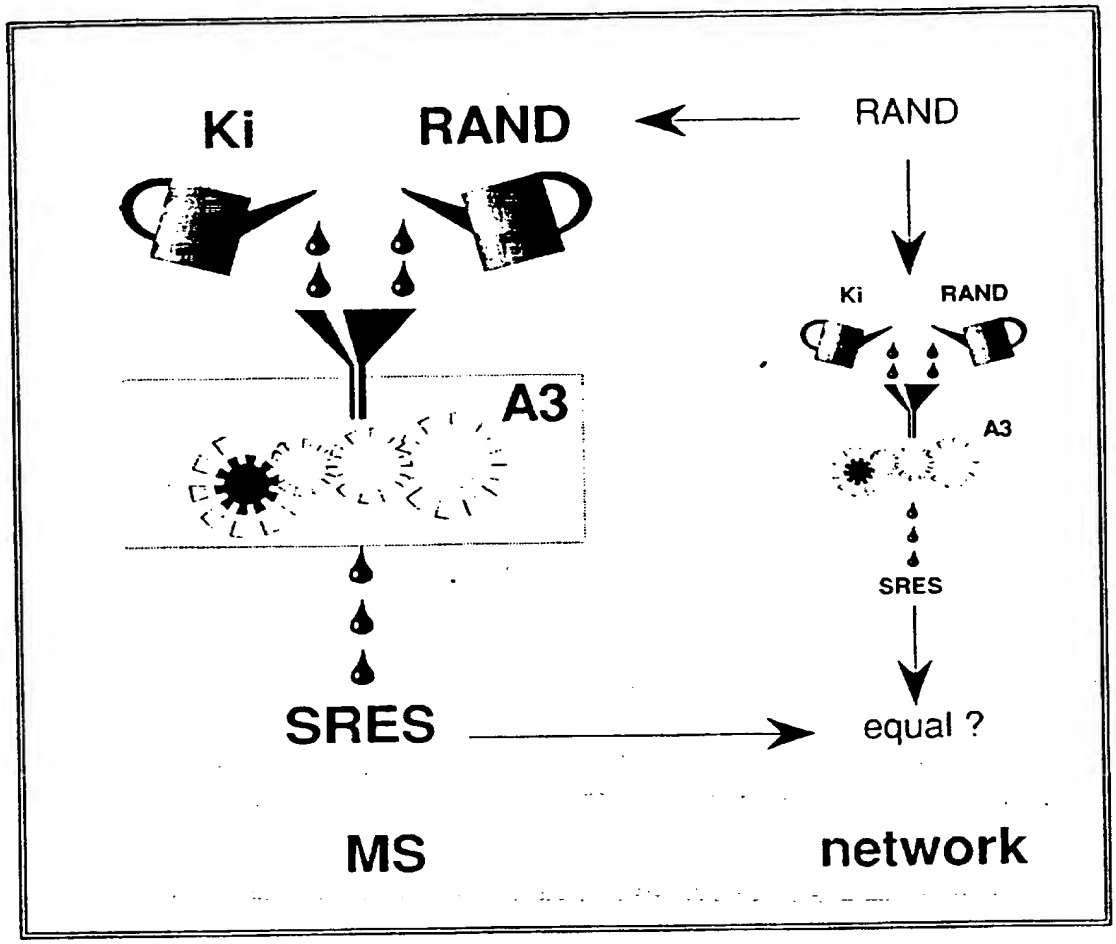


Figure 7.7 – The authentication computation

Authentication is performed by requiring the correct answer to the following riddle: what signed response *SRES* are you able to derive from the input challenge *RAND*, by applying the A3 algorithm with your personal (secret) key *Ki*?

inter-PLMN roaming. Operators can therefore choose the A3 applicable to their own subscribers independently from other operators. Such algorithms are usually kept secret (belt and braces are never too much in this domain!).

In order to obtain to desired security level, A3 should be what cryptography experts refer to as a one-way (or trap-door) function. This means that the computation of *SRES* from *Ki* and *RAND* should be easy, whereas the computation of *Ki* knowing *RAND* and *SRES* should be as complex as possible. It is indeed this level of complexity which determines which security level has been achieved. Even with the knowledge of several pairs (*RAND*, *SRES*) pertaining to the same subscriber (i.e., the same *Ki*), the computation should remain highly complex. Beyond this requirement, the only constraint imposed on A3 is the size of the input parameter (*RAND* is 128 bits long) and the size of the

output parameter (*SRES* must be 32 bits long). *Ki* can indeed be of any format and length: here again the design choices leave the operator with a maximum flexibility. Only if *Ki* would be transported in the network (see page 488) would it be constrained to a maximum length of 128 bits.

At first view, it may surprise the reader to learn of the possibility for each operator to choose A3 independently, given the general specification philosophy of GSM. Special efforts were necessary to cover the case of international roaming, where the specification of a single A3 algorithm could appear an easy solution. However, several reasons justify the approach. One of them is the administrative complexity linked to the specification and distribution of cryptographic algorithms, especially when they are to cross borders. As will be explained later on, the algorithm used for ciphering in GSM is unique, and its specifications are managed in a totally different way from the other *Specifications*. The management of a single A3 algorithm would have been even more complex, since authentication is more sensitive than communication ciphering (the consequences of a "broken" algorithm are more far-reaching in the case of authentication). The management of A3 is much simpler if controlled by a single operator. Another reason is the existence of algorithms fit for authentication and already implemented in smart cards, but possibly not open for sharing. A limiting factor being the smart card memory capacity, the choice of having an operator-dependent A3 algorithm enables telecommunication operators to use a single algorithm for, e.g., GSM SIM and pay-phone access.

7.2.2.2. Encryption

Obtaining a good protection against unauthorised listening is not an easy matter with analog transmission, but digital transmission permits an excellent level of protection with relatively simple means, thanks to digital cryptography methods. This has been taken advantage of in GSM, where the position of the encryption and deciphering processes in the transmission chain allow a single method to be used for protection of all transmitted data in dedicated mode, whether user information (speech, data, ...), user-related signalling (e.g., the messages carrying the called phone numbers) or even system-related signalling (e.g., the messages carrying radio measurement results to prepare for handover). This choice is not the result of a paranoiac approach, but is justified by its simplicity. Only two cases need to be distinguished: either transmission is protected, and everything is sent enciphered, or transmission is not protected, and everything is sent in clear text. The actual procedure for changing from

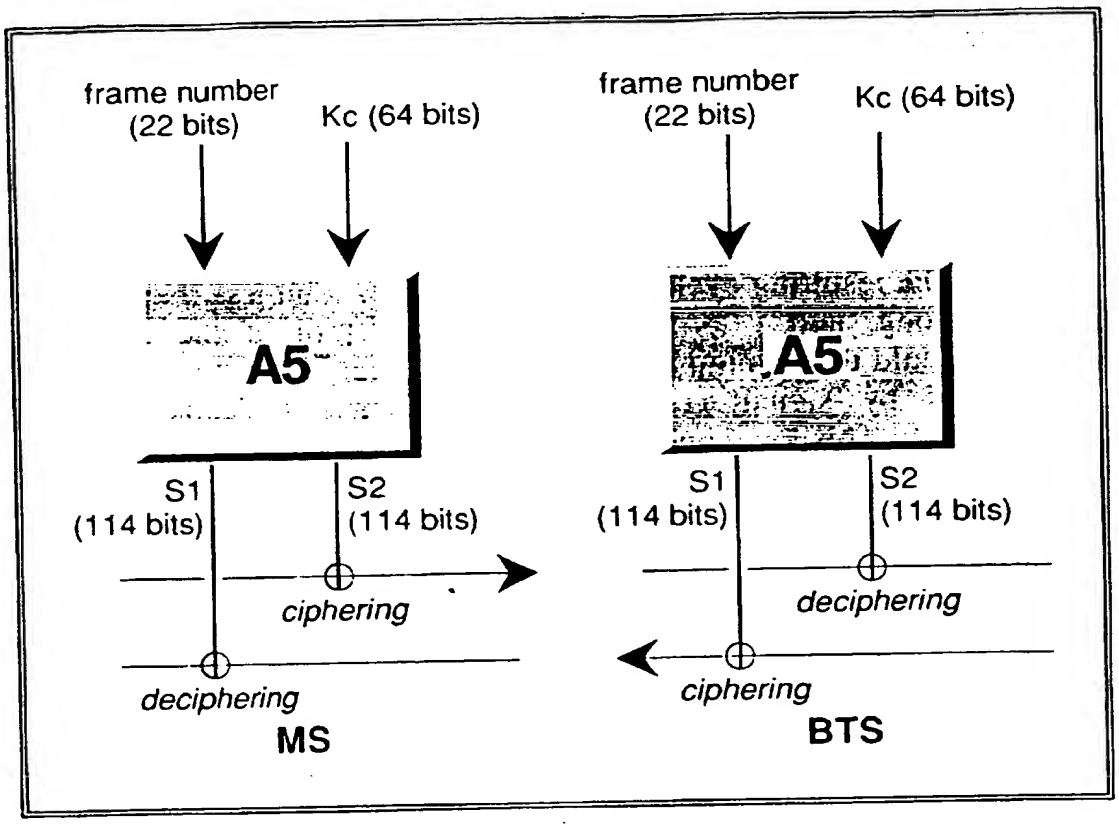


Figure 7.8 – Ciphering and deciphering

A5 derives a ciphering sequence of 114 bits for each burst independently, taking into account the frame number and the ciphering key *Kc*.

ciphered to non-ciphered mode (and vice versa) belongs to the radio resource management functions and has been described in Chapter 6.

Both ciphering and deciphering are performed by applying an "exclusive-or" operation between the 114 "coded" bits of a radio burst and a 114-bits ciphering sequence generated by a specific algorithm, called A5, as described in Chapter 4. In order to derive the ciphering sequence for each burst, A5 performs a computation with two inputs: one is the frame number and the other is a key (named *Kc*) agreed between mobile station and network (see figure 7.8). The uplink and downlink directions use two different sequences: for each burst, one sequence is used for ciphering in the mobile station and for deciphering in the BTS, whereas another one is used for ciphering in the BTS and deciphering in the mobile station.

For all types of radio channels, the frame number changes from burst to burst, so that each burst of a given communication in the same direction uses a different ciphering sequence. The successive values of the frame number depends on the time organisation of each channel. The

time organisation of a TACH/E, exposed in Chapter 4, shows for example that the frame number is not always incremented by one every burst.

As far as its representation is concerned, the frame number is coded as the concatenation of three values, called respectively T1, T3 and T2 (in that order) and amounting to 22 bits all together. What these three values actually represent is in fact meaningless as far as ciphering is concerned. The resulting cycle, the hyperframe, is a bit less than 3 and a half hours long, and determines a periodic return of the ciphering sequence, should the communication last this long. However, the key *Kc* is controlled by signalling means and changes typically at each communication. This key is not publicised, but, since it is often changed, it does not need as strong a protection as *Ki*: for instance, *Kc* can be read freely from the SIM.

Algorithm A5 must be specified at the international level, since for achieving MS-roaming it must be implemented within every base station (as well as in any mobile equipment). Means to cater for several A5 algorithms (e.g., to cope with some regulation restrictions with regards to export outside Europe) have been introduced in the *Specifications* (partially in phase 1, and definitely for phase 2). For the time being, a single A5 algorithm has been specified for use in all countries. Its specification cannot however be found in the *Specifications*, for security reasons. This algorithm is the property of the GSM MoU, and is tightly copyright protected. Its external specifications are however public, and it can be described as a black box taking a 22-bit long parameter (the frame number) and a 64-bit long parameter (*Kc*) to produce two 114-bit long sequences. As for the authentication algorithm A3, the level of protection offered by A5 is determined by the complexity of the reverse calculation, i.e., the computation of *Kc* knowing two 114-bits ciphering sequences and the frame number.

Key Management

The key *Kc* must be agreed by the mobile station and the network prior to the start of encryption. The choice in GSM is to compute the key *Kc* independently from the effective start of encryption, during the authentication process. *Kc* is then stored in a non-volatile memory inside the SIM, so as to be remembered even after a switched-off phase. This "dormant" key is also stored in the visited MSC/VLR on the network side, and is ready to be used for a start of encryption. When authentication happens while the transmission is ciphered, then the active key *Kc* being used for ciphering/deciphering is not affected, but the new "dormant" key is stored, and is reserved for use at the next occurrence of

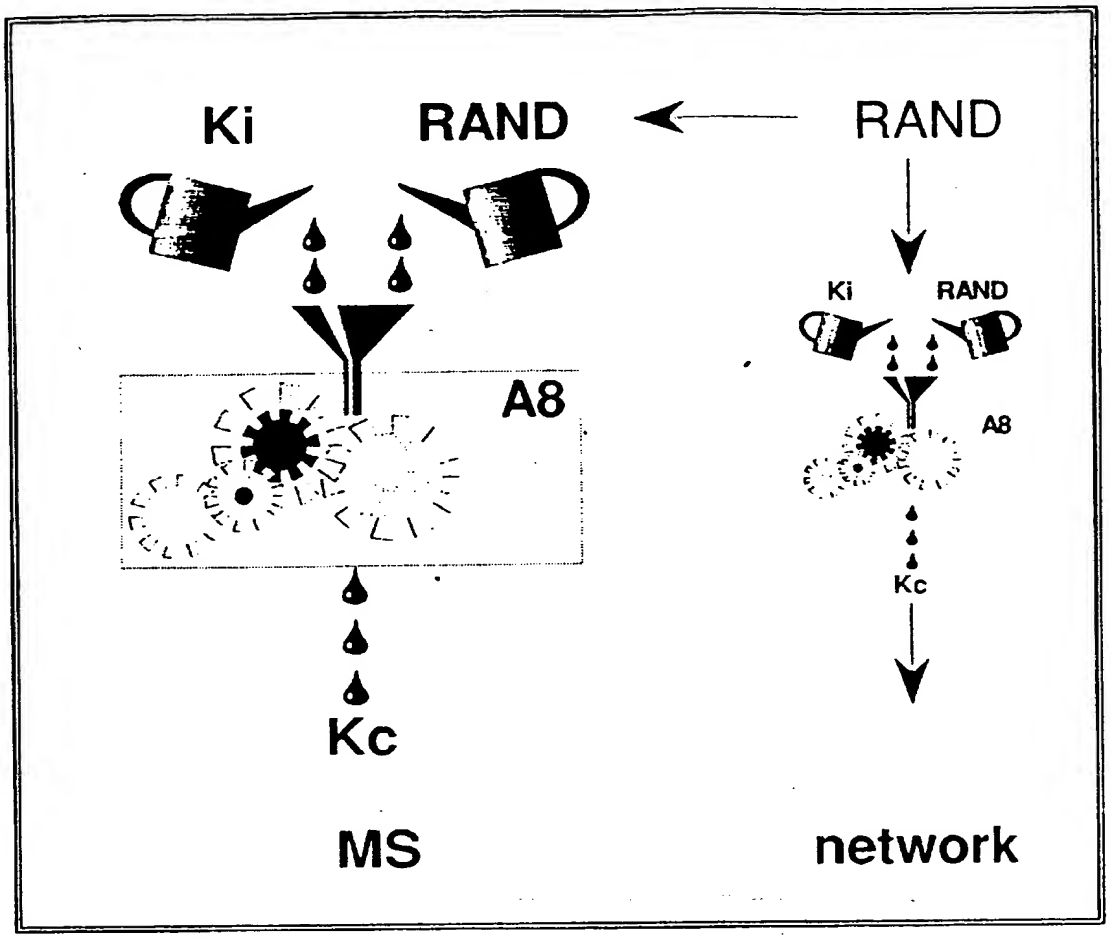


Figure 7.9 – Kc computation

Each time a mobile station is authenticated,
the mobile station and the network also compute the ciphering key Kc
by running algorithm A8 with the same inputs RAND and Ki
as for the computation of SRES through algorithm A3.

a transition between clear mode and cipher mode. Hence the terminology "dormant" key (versus "active" key) introduced in this book.

The algorithm used to compute *Kc* from *RAND* (the same one used for authentication) and *Ki* is called A8 in the *Specifications* (see figure 7.9). Similarly to A3 (the authentication algorithm computing *SRES* from *RAND* and *Ki*), A8 is not specified in the ETSI *Specifications*, but can be chosen independently by each operator. Both algorithms could in fact be implemented as a single computation. For instance, they could be implemented as a single algorithm whose output consists of 96 bits: 32 bits to form *SRES*, and 64 bits to form *Kc*. Care must be taken that the knowledge of *RAND* and *SRES* does not give away too much information about *Kc*.

It is worth noting that the length of the significant part of key *Kc* output by algorithm A8 is fixed by the group of signatories of the GSM MoU, and may be less than the maximum 64 bits. In that case, the significant bits are complemented with zeroes, so that the format always uses the full 64 bits. As far as A5 is concerned, all patterns of 64 bits are possible and meaningful: this mechanism allows the level of security to be increased in the future if needed, without any change to A5 (therefore without any change of the mobile equipments) by increasing the number of significant digits within the limit of 64.

Since A3 and A8 are always running together, and in most cases are implemented as a single algorithm, they will always be treated together in the rest of this chapter, and referred to as A3/A8. They are indeed so intermixed that authentication and "dormant" key management cannot be dealt with independently.

No other "Ax" algorithms are to be found in the Specs. A1, A2, A4, and so on, were place-holders during the design of the system, and disappeared eventually, but the terminology of the three survivors A3, A5 and A8 was not changed.

7.2.2.3. User Identity Protection

Encryption is very efficient for confidentiality, but cannot be used to protect every single exchange on the radio path. Ciphering with *Kc* applies only when the network knows the identity of the subscriber it is talking to. Obviously, ciphering cannot be applied to common channels, such as the BCCH, which is received simultaneously by all mobile stations in the cell and in neighbouring cells, or else it could be applied with a key known to all mobile stations, and therefore quite useless as a security mechanism! When a mobile station moves to a dedicated channel, there is some "bootstrap" period during which the network does not yet know the identity of the subscriber, say Charles, and therefore cannot cipher. This has a major consequence: all the signalling exchanges up to and including the first message carrying a non-ambiguous subscriber identity must be sent in clear. A third-party could at this stage listen to this identity, and know where Charles roams at this particular moment. This is considered harmful to Charles' privacy, and a specific function has been introduced in GSM to cater for such confidentiality.

Protection is obtained by using an identity alias, the TMSI (Temporary Mobile Subscriber Identity), which is used instead of the subscriber identity (the IMSI) when possible. This alias must be agreed before-hand between the mobile station and the network, during protected (ciphered) signalling procedures. Since this confidentiality feature is

totally independent of the other security functions, the description of the corresponding mechanisms will be dealt with separately from authentication/ciphering later on in this chapter.

7.2.3. ARCHITECTURE AND PROTOCOLS

The actors and protocols involved in security management are almost the same as for location management, and this justifies their inclusion in the same functional plane. However, for security management, the starring roles are displaced, and must be attributed to the SIM on the mobile station side, and the Authentication Centre (AuC), which can be seen as a part of the HLR, on the network side.

The SIM and the AuC are the repositories of the key *Ki* of the subscriber. They do not transmit these keys, but perform the A3 and A8 computations themselves. As far as authentication and setting the key *Kc* are concerned, all other involved equipments are intermediaries.

The AuC is not involved in other functions than the ones just listed, concerning the GSM radio path security management. The AuC may be implemented as a separate machine or as modules of the HLR. The main reason for the distinction between AuC and HLR in the *Specifications* is to sensitise operators and manufacturers to the security issue. As mentioned earlier, all the security mechanisms described in this chapter rely on the secrecy of *Ki*. The AuC is a means to build an additional layer of protection around the *Ki*'s.

The SIM takes responsibility for most of the security functions on the mobile station side. It stores *Ki*, it implements the operator-dependent A3/A8 and it also stores the "dormant" Kc. The existence of the SIM as a separate physical piece from the mobile equipment is indeed one of the elements enabling flexibility in the choice of A3/A8. The mobile equipment manufacturers need not be aware of the specifications of these algorithms for any operators. The SIM manufacturers, on the other hand, must implement potentially different algorithms for each of their operator-customers, but competition, mass-market production and distribution issues are totally different compared with the mobile equipment market.

The SIM protects completely *Ki* against reading. The smart card technology, introduced some time before GSM to produce these tiny electronic safes, was exactly fitting for this purpose. The only access to *Ki* happens during the initial personalisation phase of the SIM, when *Ki* is

written in the SIM. This phase happens under the tight control of the operator. Later on, *Ki* is only accessed internally within the SIM when it has to compute SRES and Kc: a procedure on the SIM-ME interface allows the mobile equipment to send a value *RAND* and to get in return, typically a few tens of milliseconds later, the corresponding *SRES* and *Kc*. Another advantage of SIM-storage for *Ki* lies with the possibility, if security requires (e.g., as a regular measure, or if it turns out that the chosen A3/A8 is not as secure as expected) to issue a new SIM on a per-subscriber basis.

The MSC/VLR plays several small roles. It initiates authentication; it decides to switch to ciphered mode; it checks the SRES provided by the SIM (through the mobile station) with the one provided by the AuC (through the HLR); it stores the "dormant" Kc on the network side; and it manages the TMSI.

Ciphering is a transmission function, and as such it involves transmission equipments (the BTS for instance), and the radio resource management protocols (and the BSC). These aspects were tackled in the respective sections, and will not be revisited here. This chapter deals only with the decision to go to ciphered mode, as well as with the management of the needed parameters.

The security management functions at this level are supported by the same protocols (plus some others) as seen for location management. The RIL3-MM protocol supports the dialogue between the mobile station and the MSC/VLR, whereas MAP/D is used between the MSC/VLR and the HLR.

The SIM-ME protocol in the area of security management uses more than just read or write commands. Others are added to provide *RAND* and to request for an A3/A8 computation, as well as for getting back *SRES*.

The additional protocols compared to those used in the location management area include the protocol between HLR and AuC, which is only indicatively specified in the *Specifications*, as part of the general Operation and Maintenance GSM protocol. Last, a small additional protocol is introduced between MSC/VLRs to enable an MSC/VLR to ask another one for the identity and subscription data of a user, before access to the HLR. This protocol is used upon access of a mobile station identified by a TMSI with reference to another MSC/VLR. It is the MAP/G protocol and is limited to one operation.

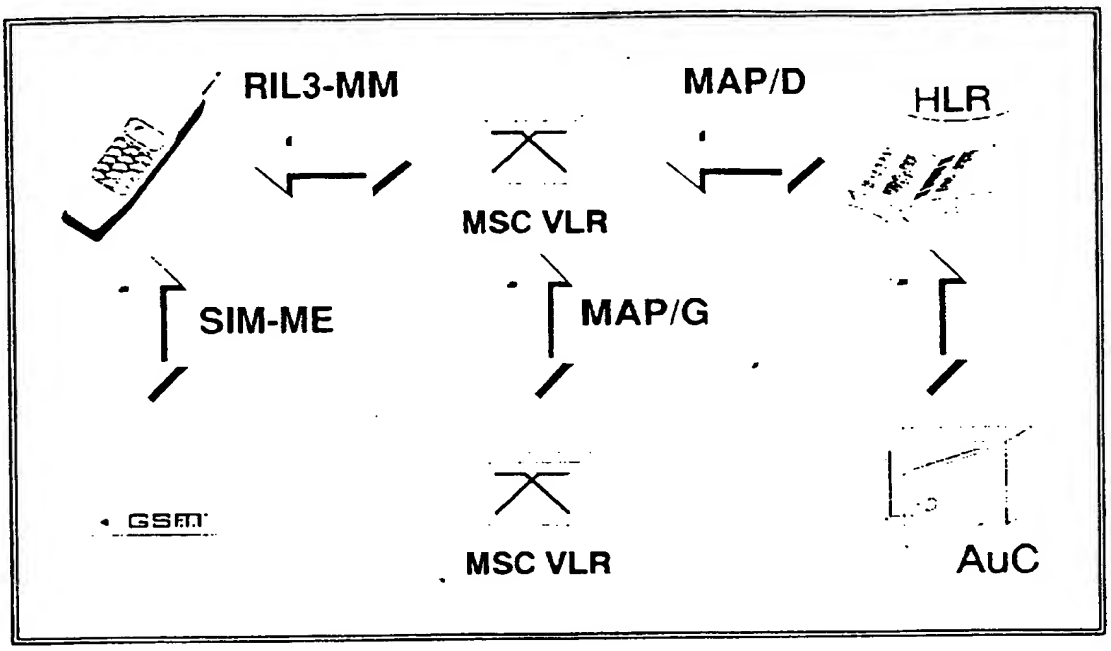


Figure 7.10 – Security management protocols

Security management is coupled with location management, and makes use of the same protocols, with two additions: a small protocol to transfer subscriber data between MSC/VLRs (MAP/G), and the connection of the AuC to the HLR.

The protocol structure is summarised in Figure 7.10.

7.2.4. THE SIGNALLING MECHANISMS

7.2.4.1. Authentication and Encryption Key Management

As explained in the preceding section, the computation of the signed response for authentication and of the ciphering key *Kc* are performed simultaneously, based on the same inputs *Ki* and *RAND*, the latter changing every time. The signalling mechanisms used for managing the corresponding data in the network are strongly coupled and will therefore be described together.

The key *Ki* is a subscriber parameter. As such, it is stored in the HLR, more precisely in the AuC. The authentication and key setting procedure, on the other hand, is controlled by the visited MSC/VLR, which decides when to run this procedure: e.g., at call set-up, location updating, etc. There are then two different procedural aspects, distinct in time as we will see: the real-time authentication and key setting procedure between the mobile station and the MSC/VLR, and the

procedure for transporting security related data between HLR/AuC and MSC/VLR.

The MS-MSC Procedure

The authentication procedure between visited MSC/VLR and mobile station consists of two messages: the RIL3-MM AUTHENTICATION REQUEST message from visited MSC/VLR to mobile station, transporting *RAND*, and the corresponding RIL3-MM AUTHENTICATION RESPONSE answer from the mobile station, giving *SRES* for checking in return.

At the reception of the RIL3-MM AUTHENTICATION REQUEST message, the mobile equipment sends to the SIM a RUN GSM ALGORITHM message, containing *RAND*, and immediately after a GET RESPONSE message whose answer contains *SRES* and *Kc*. *SRES* is sent back to the network in the RIL3-MM AUTHENTICATION RESPONSE message, whereas *Kc* is written back in the SIM at the right place, and so kept for further use.

An interesting procedural detail concerns the sequence number of *Kc*, also called CKSN (Ciphering Key Sequence Number) in the *Specifications*. To cope with possible inconsistencies between the dormant *Kc* on the infrastructure side, and the one on the mobile station side, a sequence number is associated with it. This number is provided by the MSC/VLR in the RIL3-MM AUTHENTICATION REQUEST message, and is stored together with the dormant *Kc* in the SIM and in the subscriber record in the MSC/VLR. This number is given back to the MSC/VLR in the initial message in the access procedure, so that it can be checked. If not consistent, the MSC/VLR knows that an authentication procedure is needed before ordering the ciphered mode. Value 0 of the sequence number corresponds to "no *Kc* allocated".

The MSC-HLR Procedures

The computation of *SRES* and *Kc* on the network side, requiring the knowledge of both Ki and A3/A8, must be performed so that its result is made available in the visited MSC/VLR. Two options are allowed in the *Specifications*: either the computation is done in the visited MSC/VLR, or in the AuC.

The first possibility (computation in the visited MSC/VLR) requires the visited network to cope with different A3/A8 algorithms depending on the home PLMN operator. In this scenario, the HLR has no specific role for this function, and the AuC does not exist: *Ki* is just

another subscriber parameter among many others, to be stored and provided to other equipments when perchance requested. Ay! there's the rub: this scheme implies that the key *Ki* circulates through the SS7 network, introducing a weakness in the system security, because interception cannot be precluded. Moreover, international roaming requires agreement between operators as far as A3/A8 is concerned. Either a single A3/A8 algorithm is standardised on a multilateral basis, or each operator must undertake to provide the specifications of its own A3/A8 algorithm to all others, the most cumbersome aspect being the implementation of them all in each visited MSC/VLR!

The second solution overcomes both the security breach and the roaming problem, by having the computation performed in the HLR, in the AuC precisely. No need to transfer *Ki* any more, no need to divulge A3/A8 specifications either! However, signalling means must be devised to transfer the result of the computation from HLR to the visited MSC/VLR. In order to avoid such a transfer every time the visited MSC/VLR decides that it must authenticate, the computation is done in advance. For each computation, the AuC must draw a value for *RAND* and apply A3/A8. The result is a triplet of values: (*RAND*, *SRES*, *Kc*) to be sent to the visited MSC/VLR. The visited MSC/VLR stores a reserve of a few such triplets per subscriber, in which it can draw at need. This reserve is first established when the subscriber first registers in the visited MSC/VLR: it is part of the subscriber data provided by the HLR in the MAP/D INSERT SUBSCRIBER DATA message. Tight security requires that a triplet be used only once. As a consequence, when the reserve falls below some threshold, the visited MSC/VLR asks the AuC, through the HLR, for more triplets. The only accepted exception to this "throw away after use" rule is when a communication failure has occurred between the visited MSC/VLR and the HLR. The triplet replenishing procedure consists in two messages: the MAP/D SEND PARAMETERS message and its answer, the MAP/D SEND PARAMETERS RESULT message.

7.2.4.2. User Identity Protection

The Temporary Mobile Subscriber Identity (TMSI) is an alias for the subscriber identity used in order to avoid sending the IMSI in clear on the radio path. The TMSI is allocated by the network on a location area basis; at a given moment, it refers non-ambiguously to a subscriber when used in conjunction with the location area identity (LAI). Strictly speaking, the term TMSI should be used to refer to the full digit string composed of the LAI and of the digit string allocated at a given moment to a certain mobile station in this location area (which shall be called here "TMSI-code" or TIC). However, in the *Specifications*, TMSI is more

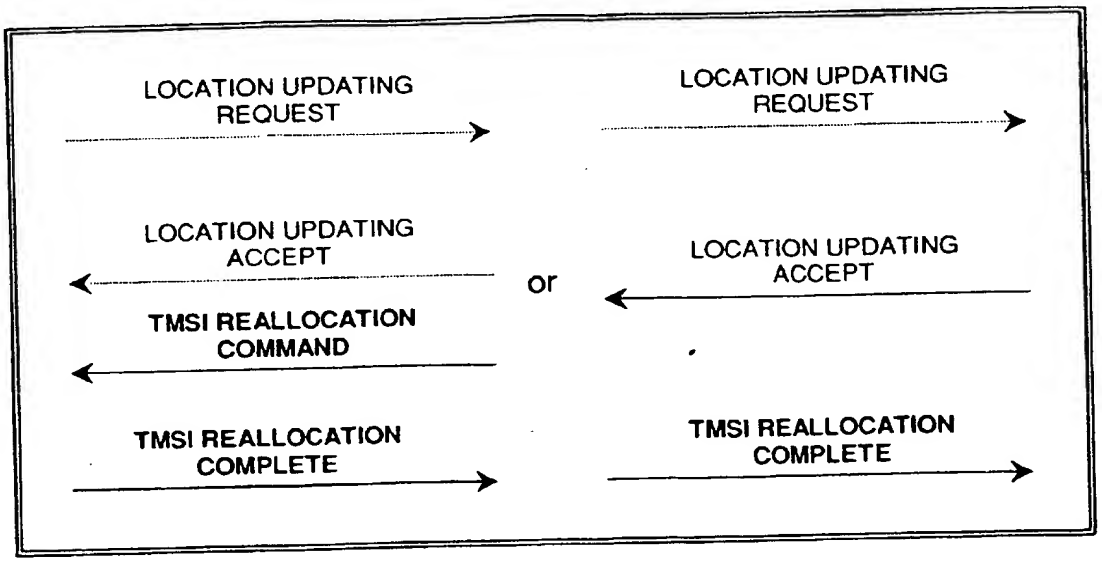


Figure 7.11 – TMSI allocation

A new TMSI can be allocated to a mobile station through a standalone procedure consisting of two messages. However, a more economical sequence is allowed when TMSI allocation is performed in conjunction with a successful location updating procedure.

often used to refer to the TIC, hence with ambiguity when the context of the location area is not clear. On the radio path, most connections are set-up in the location area in which the mobile station is registered. The TIC is in such cases sufficient to identify the subscriber non-ambiguously, the LAI being implicitly equal to the one the cell belongs to. The only exception to this rule is when the mobile station must perform a location updating attempt in a cell of a new location area: the full TMSI (including the "old" LAI) must then be used.

The length of the TIC is 4 octets, whereas the IMSI consists of up to 15 digits, coded in 9 octets including the length indicator. The short length of the TIC allows spectrum saving on the radio path when it can be used alone. This is especially the case for paging messages, where more than twice as many mobile stations may be paged in a single message when using the TIC, which cannot be ambiguous in these messages.

But the most interesting area of study concerns the management of TMSIs, i.e., how and when they are allocated and released. In the network, TMSIs are managed by the MSC/VLR (by the VLR in the full-blown canonical architecture). A TMSI is first allocated to a mobile station the first time it registers in the location area, and is released when the mobile station leaves it. A standalone allocation procedure uses two messages: the RIL3-MM TMSI REALLOCATION COMMAND message from MSC/VLR to mobile station and the corresponding RIL3-MM TMSI

REALLOCATION COMPLETE acknowledgement. However, when the TMSI allocation is performed just after a successful location updating (which is usually the case), the allocation message can be "combined" with the RIL3-MM LOCATION UPDATING ACCEPT message from the network: the new TMSI is then part of the RIL3-MM LOCATION UPDATING ACCEPT message and the RIL3-MM TMSI REALLOCATION COMMAND is dispensed of, but not the acknowledgement. The full sequence of 4 messages is however allowed as well. The corresponding sequences are shown in figure 7.11.

TMSI cancellation is usually implicit. In the mobile station, cancellation is automatic upon allocation of a new TMSI, or upon location updating acceptance in a new location area. Explicit cancellation can also be done, by sending the IMSI in the RIL3-MM LOCATION UPDATING ACCEPT message: this is to be understood by the mobile station as a cancellation of the previous TMSI.

Another feature of the TMSI allocation procedure enables the network to use a short version of the RIL3-MM TMSI REALLOCATION COMMAND message in order to allocate a TMSI having the same TIC part (the 4 octets meaningful within a location area) as the previous one. This is achieved by not mentioning any identity in the RIL3-MM allocation message (whether TMSI REALLOCATION COMMAND or LOCATION UPDATING ACCEPT). The gain brought by this feature seems however negligible.

Table 7.1 shows a summary of how to interpret the allocation message depending on its contents as far as the *MOBILE IDENTITY* parameter is concerned.

The TMSI is stored in the subscriber's record held by the MSC/VLR, but not in the HLR. When the record is destroyed, for example upon location cancellation from the HLR, the TMSI is then cancelled implicitly in the infrastructure. This raises some problems when database failures are considered. For instance, some situations may result

Identity used in the message	New TMSI
none	LAI (new) + TIC (old)
TIC (new)	LAI (new) + TIC (new)
IMSI	none

Table 7.1 – Rules for deriving a new TMSI upon allocation

Depending on the contents of the message sent by the network, a mobile station can derive the new value—if any—of its TMSI.

in the mobile station using a TMSI which is not allocated to its subscriber any more, or worse which is allocated to another subscriber. In order to cope with such situations, some tools have been included in the *Specifications*.

In particular, a procedure allows the network to ask the mobile station for its full IMSI. Typical usage of such a procedure is when the TMSI by which the mobile station identifies itself is not known in the MSC/VLR, or, if known, after the failure of authentication, which may reflect a TMSI discrepancy between mobile station and network. The identification procedure consists of two messages, the RIL3-MM IDENTITY REQUEST message sent by the network and the corresponding RIL3-MM IDENTITY RESPONSE answer by the mobile station. The procedure is in fact more general than a request for the IMSI, since other types of identity may be asked for.

When a database failure has occurred in the MSC/VLR, all the TMSIs stored in the salvaged records are dubious. In this situation, when an incoming call arrives, the MSC/VLR pages the mobile station throughout its area (instead of just in one location area) using the IMSI instead of the TMSI.

A side effect of the TMSI is its usage in an RIL3-MM LOCATION UPDATING REQUEST message sent in a location area managed by an MSC/VLR other than the one which allocated the TMSI. We have seen that in this case the mobile station gives the full TMSI, and not only the TIC as for a call setup or a response to paging. While the full TMSI is unambiguous, it does not indicate the HLR or even the home PLMN as the IMSI does. But it does give the indication of the MSC/VLR which allocated the TMSI, and which then presumably knows the corresponding IMSI.

The new MSC/VLR has two possibilities. The first is to ask the mobile station for its IMSI. This is simple, but this is a breach in the protection of the subscriber identity. The other possibility is to request the IMSI from the previous MSC/VLR, as indicated by the LAI part of the TMSI. This is supported by the MAP/G procedure (the only one in this protocol), consisting of a MAP/G SEND PARAMETERS message and the answer, the MAP/G SEND PARAMETERS RESULT message. The first message includes the list of the desired data, i.e., the IMSI and optionally authentication triplets. The answering message includes the requested data. An authentication triplet allows the new MSC/VLR to perform the authentication before the HLR answer, in order to gain time.

7.3. MISCELLANEOUS MM FUNCTIONS

As explained at the beginning of the chapter, the Mobility Management realm includes more than just location management and security management, at least if all the procedures in the RIL3-MM protocol are taken into account. Even though this chapter might be a debatable location for some of these other functions, it was felt best to keep their description here, so as not to wander too far away from the *Specifications* description. It should not be forgotten that the goal of a structured protocol modelling is intended as a help for understanding more than as an implementation guideline. A protocol architecture can therefore be to a great extent subjective.

There are four MM miscellaneous functions not yet described. All of them concern only the mobile station and the MSC/VLR. In fact only two of them involve procedural exchanges (in the RIL3-MM protocol), the two remaining ones being specifications of the mobile station behaviour. These procedures or specifications are related in so far as they use a common modelling concept. the MM-connection. This should not be mixed with the concept of CM-transaction (CM for Communication Management), which refers to a transaction in the upper layer, the Call Control plane. A CM-transaction corresponds to a call transaction (RIL3-CC protocol), to a Short Message transaction, or to a Supplementary Service management session. A CM-transaction corresponds then to all the activities described in the next chapter. After the presentation of these functions, we will discuss how useful this modelling is.

Generic Mobile Originating CM-Transaction Establishment

In ISDN call set-up procedures, the first message on the originating access interface is the SETUP message. This message contains a lot of information, including the calling number. In GSM, privacy on the radio path requests that this first message be ciphered. However, because the decision to cipher lies with the infrastructure, a preliminary message from the mobile station is necessary to give enough information to the network for it to decide whether to apply ciphering (and authentication) or not. An alternative would have been to cipher systematically.

Because this preliminary message does not exist in ISDN, because of the will to keep call control as little adulterated as possible by the specific aspects of radio transmission, and because the same need exists for other kinds of services such as short message transfer or

supplementary services, a generic mobile originating establishment procedure was introduced as part of the RIL3-MM protocol. When initiated by the mobile station, a call, a short message session or a supplementary service management session all must use the generic establishment procedure, even if transmission means are already established and used in ciphered mode. This allows the infrastructure to perform authentication, and/or to go to the ciphered mode before any further progress of the session.

There is no equivalent to the mobile station-originated generic establishment procedure if the session is initiated from the network side, simply because the network chooses to apply authentication and/or ciphering before starting any upper layer procedure.

The generic mobile station originated establishment procedure consists basically in a preliminary message sent by the mobile station, the RIL3-MM CM SERVICE REQUEST message, and a signalling sequence in answer from the MSC. The RIL3-MM CM SERVICE REQUEST message may be an initial message, as described in Chapter 6. The reactions of the MSC can be to start an authentication procedure (RIL3-MM AUTHENTICATION REQUEST message), or to answer positively the request. This can be done by sending an RIL3-MM CM SERVICE ACCEPT message, which seems normal, or by starting a ciphering mode setting procedure (RIL3-RR CIPHERING MODE COMMAND message), which is a curious specification. Still another possibility for the MSC is to reject the request by sending an RIL3-MM CM SERVICE REJECT message. If the answer is positive, the mobile station may start to initiate the procedure which justified the previous actions, for instance by sending an RIL3-CC SETUP message.

The only reason for "mixing up" the generic establishment procedure with the cipher mode setting procedure (from another protocol!) lies in the will to reduce the number of messages. A shortcoming of this method is however the ambiguity thus introduced: there is no means to distinguish an RIL3-RR CIPHERING MODE COMMAND message acknowledging a CM-transaction establishment and one which does not. "Collision" cases may well occur where the MSC wishes to start a ciphered session at the same time as the mobile station initiates a CM-transaction for another purpose. This situation will likely be improved in phase 2.

Another important point to note is the lack of connection reference in an RIL3-MM CM SERVICE REQUEST message and in the corresponding acknowledgement. This leads to ambiguity if two generic establishment

procedures are run in parallel. Hence, this is forbidden by the *Specifications*. We will come back to these issues at the end of the chapter.

Upper Layer Synchronisation

The *Specifications* require that CM-transactions cannot be initiated while a location updating procedure is running. The requirement is even more stringent in GSM phase 1: the mobile station must go back to idle mode before setting up a new RR-connection aimed at supporting a CM-transaction.

This requirement arises from the need for a subscriber to be correctly registered with the network before accessing any service. In other words, the rationale requires that no other Mobility Management or Call control procedure be started during a location updating procedure in a location area different from the one in which the mobile station was previously registered, at least up to the reception of the RIL3-MM LOCATION UPDATING ACCEPT message. An alternative would have been to allow the mobile station to anticipate the closing of the location updating procedure and to require the MSC/VLR to store the service request until (and if) location updating is successful with the HLR. This is not allowed in the phase 1 *Specifications*. A milder position could have been to allow the mobile station to send a request just after the reception of the RIL3-MM LOCATION UPDATING ACCEPT message, or to anticipate in the case of a periodic location updating. But even this is forbidden in phase 1. The only reason for such a drastic approach was the simplicity of the MSC/VLR.

Whatever the shortcomings, the specification must be implemented as it is. The impact of this synchronisation function, modelled in the MM plane in the *Specifications*, lies only with the mobile station and does not involve any procedure.

Infrastructure Activity Monitoring

A third function modelled within the MM plane and also not involving any protocol procedure is a watchdog for MSC signalling activity. Radio channel release is indeed a privilege of the infrastructure. In case of failure, the mobile station may then find itself in a difficult situation, with an unused dedicated channel which it is not allowed to

release explicitly. In such cases, the mobile station must go back to idle mode autonomously. This requires that the mobile station continuously checks if there is—to its knowledge—a CM-transaction in progress. In the opposite case, the mobile station waits some time and decides to go back to idle mode if nothing happened in-between, without sending any message to the network. The corresponding watchdog function is modelled in the *Specifications* through a timer called T3240. This timer could just as well be part of the RR plane, but is specified in the MM plane of the *Specifications*.

Re-Establishment

When a mobile station, being provided with some service, suddenly looses contact with the infrastructure, there is a possibility to resume this contact, for instance in another cell. Such cases may happen for example in configurations where the handover procedure proves to be too slow. A salvaging attempt by the mobile station has been introduced in the *Specifications*, and has been modelled in the MM plane. This is called the re-establishment procedure. All mobile stations must support this procedure, but it is optional on the network side. If it is supported, call contexts must be kept a little while after the contact loss, to allow potential re-establishment to be effective.

This feature is very close to what is called mobile station-triggered handover in other systems. It fulfils the same requirement as handover, but in a much less controlled way, though with possibly a better efficiency in configurations where propagation loss is very steep. Because of this analogy, the re-establishment procedure has been studied in Chapter 6. The initial message to request a re-establishment is the RIL3-MM CM RE-ESTABLISHMENT REQUEST. It is then worth noting that the acceptance and the rejection by the network uses the same message as for the generic CM-establishment procedure: respectively the RIL3-MM CM SERVICE ACCEPT message and the RIL3-MM CM SERVICE REJECT. The re-establishment procedure is then deeply entangled with the generic CM-transaction establishment procedure, a point which renders difficult the evolution of the re-establishment procedure.

Modelling

The four miscellaneous functions described above may seem quite ill-assorted from an architectural point of view. The two last ones would fit better in the Radio Resource plane for instance. There is however a common denominator between the four functions in the *Specifications*:

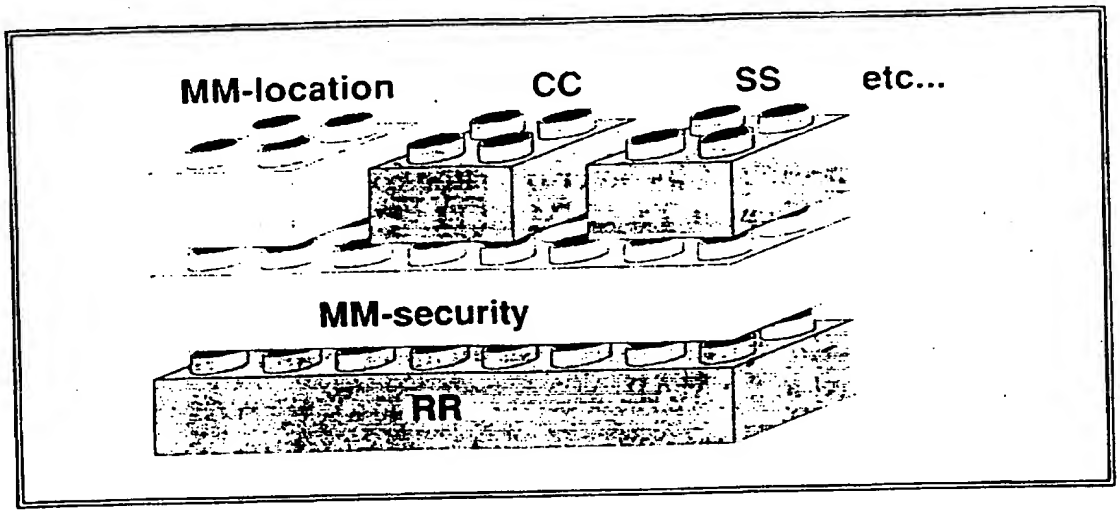


Figure 7.12 – Alternative MM modelling

GSM MM functions could be modelled in two blocks: a location management block on the same level as call control, etc. and an intermediate layer providing service to the location management block as well as to other upper layer blocks.

the concept of *MM-connection*. This pure modelling concept was not used here because it is not necessary and does not in fact help understanding. In fact the notion of MM-connection does not appear concretely in the protocols. There is a one-to-one correspondence between CM-transactions and MM-connections, which renders useless the insertion of a CM-connection identifier in the messages. MM-connections are explicitly established only when initiated from the mobile station side (by the generic Mobile Originating establishment presented above), but not when initiated from the MSC (the establishment is implicitly done by the establishment of the CM-transaction). Moreover it is never explicitly released.

An alternative to the modelling approach used in the *Specifications*, possibly reflecting the different roles of the MM plane more accurately, would consist in considering the MM functions in two separate groups (see figure 7.12). On one side, a location management functional block which would stand in parallel and at the same level as call control, supplementary service management, etc. A location updating connection would then be considered the same way as a CM-transaction. Below such blocks, and above the RR layer, a second functional MM block would provide security related services to the upper layers, namely authentication, ciphering key management, as well as the generic Mobile Originating establishment procedure and a synchronisation function whose sole aim is to forbid the establishment of upper layer procedures while a location management connection is in progress.

SPECIFICATIONS REFERENCE

The problems raised by inter-operator roaming are exposed from the service point of view in **TS GSM 02.11**. The main part of this short document concerns the choice of PLMN.

The technical aspects of location updating in the NSS are the subject of **TS GSM 03.12**. This specification mainly introduces the procedures included in the MAP.

It is worth noting that the general aspects of location updating, PLMN and cell selection in the mobile station will be presented in general terms in the future TS GSM 03.22, unfortunately not present in the phase 1 *Specifications*. Though requiring some caution, because of the functional differences between phase 1 and phase 2, this document can be useful.

The details of the cell choice algorithms, including measurement considerations are addressed in **TS GSM 05.08**, section 6.

The signalling aspects of the MM protocol between the mobile station and the infrastructure (dealing with all the topics addressed in this chapter, location updating as well as security management) are dealt with in **TS GSM 04.08**, section 4.

A very good synthesis of the general scheme for security management can be found in **TS GSM 03.20**.

The MAP protocols in general are the subject of **TS GSM 09.02**. Of relevance for this chapter, one can cite section 5.2 (location registration/cancellation), section 5.8 (fault recovery of location registers, a subject introduced in **TS GSM 03.07**), section 5.11 (management of security related functions), and section 5.15 (paging and search procedures).